DOCUMENT RESUME

ED 067 958 48 FL 003 661

AUTHOR	Hutchins, John A.
TITLE	An Investigation of Spoken Brazilian Portuguese: Part I, Technical Report. Final Report.
INSTITUTION	Naval Inst., Annapolis, Md.
SPONS AGENCY	Institute of International Studies (DHEW/OE), Washington, D.C.
BUREAU NO	BR-8-0130
PUB DATE	Aug 72
CONTRACT	OEC-0-8-000130-3543-014
NOTE	79p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Computational Linguistics; *Computers; Data Bases; Educational Experiments; *Language Research; Modern Languages; Optical Scanners; *Portuguese; Romance Languages; Speech; *Syntax; *Word Frequency

ABSTRACT

This final report of a study which developed a working corpus of spoken and written Portuguese from which syntactical studies could be conducted includes computer-processed data on which the findings and analysis are based. A data base, obtained by taping some 487 conversations between Brazil and the United States, serves as the corpus from which a frequency list of some 2,000 words is derived. A print-out of the Key-Word-in-Context is also developed and intended for use by linguistic researchers. Descriptions of experimental procedures, findings, and recommendations are included. Supportive technical data and experimental information are appended. (RL)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 001 730

FINAL REPORT

Contract No. OEC-0-8-000130-3543 (014)

AN INVESTIGATION OF SPOKEN BRAZILIAN PORTUGUESE

PART 1 - TECHNICAL REPORT

John A. Hutchins
Professor

U.S. Naval Academy
Annapolis, Md. 21402

August 1972

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare, under provisions of Title VI, Section 602 of the National Defense Education Act, Public Law, 85-864, as amended.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Institute of International Studies

FILMED FROM BEST AVAILABLE COPY

FL 003 661

FINAL REPORT

Contract No. OEC-0-8-000130-3543 (014)

U.S. Naval Institute
Annapolis, Md. 21402

AN INVESTIGATION OF SPOKEN BRAZILIAN PORTUGUESE

August 1972

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Institute of International Studies

CONTENTS

		Page
PREFACE		v
I.	INTRODUCTION	1
	Problems Under Consideration	1
	Previous Studies	1
	Objective	3
	Methods	3
	Description	4
	Transcription of Recordings	5
	Protocol	5
	Typing for Optical Scanning	6
II.	FINDINGS AND ANALYSIS	
	Description of Materials Produced	7
	Nouns	9
	Others	9
	Definite Articles	10
	Variants	11
	Description of Verbal Forms	12
	Totals of Verbal Forms	13
III.	CONCLUSIONS AND RECOMENDATIONS	19
	ENDNOTES	21
APPENDIX I	VERB FREQUENCIES	22
APPENDIX II	VERB FORMS (FREQUENCIES)	24
APPENDIX III	DISTRIBUTION OF INFORMANTS BY PROFESSION	27
APPENDIX IV	AUTHORS AND SELECTIONS LIST	28
APPENDIX V	INDENTIFICATION TAGS AND CODES	32
APPENDIX VI	SPECIAL PROVISIONS FOR TYPING LITERARY PORTUGUESE	39

APPENDIX VII COMPUTER PROGRAMS

Sample Pages for Optical Scanning . . . 40
Example of KWIC Print-out 42
Utility Programs for Languages 43
Key-Word-in-Context Program 53
Alphabetizing Letters 55
Sample Input from Teletypewritter Terminal 56
Input Verification Program 57
Forward Concordance Program 58
Forward Listing Program 64
Reverse Concordance Program 65
Reverse Listing Program 70
KWIC Print-out (Reduced size) 71

APPENDIX VIII SPOKEN BRAZILIAN PORTUGUESE FREQUENCY LIST*

a. Descending Order of Frequency
b. Alphabetical Order

APPENDIX IX LITERARY BRAZILIAN PORTUGUESE FREQUENCY LIST*

a. Descending Order of Frequency
b. Alphabetical Order

* In separate volumes which may be requested from the Project Director, Dr. J. A. Hutchins, for examination.

PREFACE

This final report is divided into two parts. Part I is the descriptive background, the frequency lists for both the spoken and the written corpus, and the most important findings together with recommendations. Part II is a study (a doctoral dissertation) by Clēa Rameh entitled, "Toward a Computerized Syntactic Analysis of Portuguese." This study was based on a segment of the spoken corpus.

The authors are fully aware that only the highlights and only a preliminary analysis of the linguistic phenomena of spoken Brazilian Portuguese can be presented here. However, we have captured an excellent sampling of the spoken language, processed it, and made it available to others. Even from a time span, it should have value in the years to come since easy retrieval of its elements is possible in a number of different ways.

The frequency lists, computer programs, and the protocol will be found in the appendix. It is hoped that microfiche and/or microfilm copies of the corrected Key-Word-in-Context lists will shortly be available for interested scholars. Also, it may be possible for the Naval Academy to put the lists "on line" in our Time-Sharing system so that users may access them over long-distance telephone lines. It is obvious that there is much to be done and that this project only represents a step in the direction of analysis of the spoken language.

There were many individuals who were instrumental in helping us over the rough spots as well as making substantial contributions. William L. Higgins, former project officer, accompanied every detail of the project and was untiring in his effort to see that the project was successful. Commander R.T.E. Bowler, Jr., USN (Ret.), provided us with an efficient and smooth fiscal operation through the U.S. Naval Institute. Professor John D. Yarbro, Chairman of the Naval Academy's Area-Languages Studies Department guided us through the difficult phase of proposal writing and setting up the project at the Naval Academy. Dr. James Nielson of the National Security Agency was most helpful with suggestions as to procedure as were Charles Holt and Marvin Peacock in the computer programming. In this same field, Dr. Harold Kaplan, Professor of Mathematics at the Naval Academy, designed some highly sophisticated computer programs which provide for upper and lower case, various accent marks, and other characters to be alphabetized in any order whatsoever once the order has been so indicated.

Yara Telles transcribed the recordings that produced our spoken corpus and prepared both manuscripts for optical scanning. Finally, my two colleagues, Dr. Guy J. Riccio, Head of the Spanish-Portuguese Division of the U.S. Naval Academy and Dr. Cléa Rameh, of Georgetown University's School of Languages and Linguistics, because of their dedication and effort, deserve much of the credit for the successful conclusion of this undertaking.

INTRODUCTION

I. Problems under consideration.

Portuguese, in spite of being one of the most widely spoken languages in the world, has long been neglected by linguists and researchers. In Brazil alone almost one hundred million persons speak Portuguese. This study is mainly concerned with the spoken language of Brazil. And as Brazil increases in importance so does its language.

Traditionally most linguistic research has been concerned with aspects of the written language. A glance at the titles of the many doctoral dissertations shows how few deal with the spoken language in spite of the fact that from ninety to ninety-five percent of all communication is transmitted in an oral form. The difficulty has always been that of obtaining a suitable spoken corpus and converting it to a non-volatile state so that the linguistic elements will be available for investigation.

Up to now no analysis of spoken Brazilian Portuguese has been attempted by scientific, empirical methods. Serious technical problems are probably responsible for the few studies completed and these are only in English, French, and German. According to H.A. Gleason - "A written language is typically a reflection, independent in only limited ways, of the spoken language. As a picture of actual speech, it is inevitably imperfect and incompete."[1] Gleason adds later that "Linguistics must start with thorough investigation of spoken language before it proceeds to study written language."[2]

Previous studies.

The only comprehensive study of Brazilian speech is that of Professor Earl Thomas of Vanderbilt, The Syntax of Spoken Brazilian Portuguese.[3] Based on a collection of notes taken over a twenty-year period, Professor Thomas was able to present a concise analysis based on his subject data. His monumental work is extremely important to scholars in Portuguese and must be taken into serious consideration. Professor Thomas did not, however, use a data base of tape recordings and statistical compilations to arrive at his conclusions. It is perhaps at this point that this study hopes to make a contribution.

Fred Ellison, at the University of Texas, started to collect and has a small corpus of spoken Brazilian Portuguese - a corpus which was obtained from directed, tape-recorded interviews with Brazilian exchange students. The interviews have been transcribed and apparently are available to scholars. Seemingly, such interviews would not fall in the category of free, natural language. The fact that those being interviewed knew they were being recorded probably would have inhibited them in some manner. Then there is also the thought that directed interviews might produce responses that lacked authenticity in natural speech patterns.

An attempt to make a survey of the spoken language of Brazil, building a corpus of over a million words of spoken Brazilian Portuguese, by a group headed by Professor Adriano Kury of the University of Brasília, was given up because of a lack of funds. While describing a research project that he was conducting on a syntactical analysis of Brazilian Portuguese, Professor Henry Hoge, Florida State University, told me that he chose the works of 27 different authors because of their colloquial style. He divided his corpus into two groups - one narrative and the other spoken or dialogue. Professor Hoge stated that "in no case was it spoken language but (it) was a representation," adding that "not much could be done with the spoken language."[4] Samples from the 27 authors selected by Hoge form the data base for our literary Brazilian Portuguese corpus.

This author contends that the "representations" of spoken Portuguese, or any other language for that matter, when found in a written form, will vary considerably from the actual spoken language. Charles F. Hockett felt that "no writing system has ever provided for the graphic representation of everything that counts (morphemically or phonemically) in speech" . . . "We do not write English as we speak it."[5] And in Computational Analysis of Present-day American English, Mary Lois Marchworth and Laura M. Bell state: "It would be well to remember that in the fiction selections of the Corpus, dialogue represents artistic rather than actual rendering of the spoken language.[6]

Written or literary Portuguese has been studied for some time. Back in 1945 Charles B. Brown, Wesley M. Carr, and Milton L. Shane compiled a Graded Word Book of Brazilian Portuguese. An "eyeball" count of 1,200,000 running words produced a total of some 26,278 different words. Of these different words only 9,345 were found in five or more different sources. The data base was extracted from prose, drama, newspapers, and technical journals. To simplify the operation, the most common (in their judgement) 222 words were excluded from the counting.[7]

In 1951 Charles Brown and Milton Shane published their Brazilian Portuguese Idiom List, containing some 3,500 items. Interestingly enough, the authors felt that "only in prose can one find the bases for normal idiomatic usage."[8] In their selections from novels and children's books, they examined only pages containing dialogue, claiming that "with this procedure it may be said that more than 50 per cent of our materials represent conversation, that is, to the extent that the printed page reflects this usage."[9]

More recently Professor John R. Kelly (Santa Barbara) developed a list of the five hundred most common words he found in a 127,000 word corpus taken from Brazilian novels, periodicals, and newspapers. Using computational methods, Kelly was able to include even the most common words in his count.[10] Also, at Stanford, John C. Duncan produced a frequency list derived from selections of continental Portuguese written between 1918 and 1939.[11]

Objective.

Our objective was to develop a working corpus, first of spoken and then written Portuguese from which syntactical studies could be conducted. The original objective was limited to establishing a list by frequency of occurrence of the 2,000 most commonly used words in spoken Brazilian Portuguese. As the project progressed, it became evident that in addition to the frequency lists, and with some increased effort, we could obtain magnificent print-outs of the Key-Word-in-Context (KWIC) for both corpora, lists which would have immense value for in-depth linguistic studies.

Methods.

There is really no ideal way to capture the spoken language, the uninhibited, natural, and spontaneous language. Various experiments and attempts have been made using such techniques as hidden microphones, direct interrogation, and recording radio broadcasts - all with less than satisfactory results. Because of this a unique method was developed, which, while it may have certain shortcomings, does give an almost natural speech pattern between two natives in a real situation in which actual information is exchanged.

The data base was obtained by taping "phone-patch" conversations between Brazil and the United States, dialogues transmitted on Amateur frequencies which are in the public domain and which can be monitored by anyone. All these conversations were recorded without the knowledge of the participants. One factor that detracted from the naturalness of the dialogues was that it was necessary to say "câmbio" or "over." However, this was more than compensated by the fact that no conversations were superimposed on others - that is,

there were never two people talking at the same time. A perfect identification could be established at all times as to who was speaking. The participants were so intensely interested in the subjects being discussed that they paid little attention to the medium being used or to their manner of speaking. Each conversation represented a unique opportunity to speak to a loved one or friend some five thousand miles away. In summation, it is quite remarkable to listen to the tapes and note the naturalness in the way the "informants" talk to each other.

Actual recording began in July 1967, being supported through a small grant from the U.S. Naval Academy Research Council. In the spring of 1968 a contract was negotiated between the U.S. Office of Education and the U.S. Naval Institute, acting in behalf of the Naval Academy, to broaden the scope of the original project so that meaningful results could be obtained concerning the speech of Brazil.

In processing the data for computer input, certain innovations were tried and proven successful. It became apparent that it would be quite feasible to handle a large size literary corpus of Brazilian Portuguese, using the same methods with only slight changes. A revision to the original contract provided for the processing of 10,000 word segments each from 27 contemporary Brazilian authors. The list of authors, titles, and selections is included in the Appendix. A few changes were incorporated in the protocol in order to transcribe the original texts as closely as possible.

Description.

Some 487 different conversations between native Brazilians were recorded on 142 reels of 600 foot tape and 34 reels of 1,200 foot magnetic tape, 7 1/2 ips., full track. Since the frequency range was from 500 to 3,000 cycles, no provisons have been made for using these tapes for phonological studies. There are a total of 837 persons, 367 men and 470 women of which at least fifteen have since passed away. About three-fourths of the men could be identified by profession, this being much more difficult for the women.

Recording was done on a random basis without a predetermined plan. A scientific sampling such as described by Leslie Kish,[12] could hardly be justified because of the tremendous difficulties, effort, and expense involved. The problem of obtaining speech samples from the many vast areas of the country is itself a formidable undertaking.

We were most fortunate in having the majority of our conversations voiced by persons living in or having come from Rio de Janeiro. The carioca accent, as Earl Thomas points out,[13] is the prestige speech of Brazil and carries with it the mystique of the cidade maravilhosa. Also, with

nationwide television programs originating in Rio de Janeiro, this accent is rapidly becoming the standard of the country. To a lesser extent we do have many dialogues originating in São Paulo, Pôrto Alegre, Curitiba, Vitória, Bahia, Recife, Belo Horizonte, Fortaleza, and other parts of Brazil. Most of the "informants" belong to the upper-middle class, have traveled extensively, and seldom reside in the place of their birth. Only a fraction show any regional speech characteristics. In comparision with a literary corpus, which we also processed, there is relatively little slang and few taboo words are present.

Many of the dialogues deal with travel, state of health, and lack of correspondence. There are several interesting little stories, such as: an air-rescue search in the Amazon, a conversation between a beauty queen and her fiancé, a description of serum for organ transplants, purchases of furniture, arrangements for scholarships, and a request for articles for "macumba" sessions. There are also discussions of arrangements for international conferences, internal administration of governmental organizations, arrangements for transmission of radio broadcasts, weather, various descriptions of American schools, several informal conversations ("bate-papos"), automobile accidents, deaths and funeral arrangements, discussions of legal affairs, military service, births, sickness and operations. A few of the dialogues refer to football games and to the purchase of clothing as well as to getting goods through customs and shipping articles to Brazil. There are several exchanges from Ric Grande do Sul which have the tu form of the verb. In general, the conversations contain samples of many of the common subjects that are usually discussed in the language.

Transcription of recordings.

In transcribing the recordings, every effort was made to reproduce as accurately as possible what was actually said. It was decided to recognize as words abbreviated forms that are frequent in speech. Examples are: cê, nê, tou, tá, and others. Because of obvious limitations in the frequency range of the reproductions, we did not attempt to determine the existence or omission of final s or z. The final "hard copy" was corrected by again listening to the original recording. All pauses, hesitations, repetitions, and false starts were included and unintelligible sounds were so indicated.

Protocol.

Before any data could be processed for computer input, it was necessary to select which symbols we would use for accent marks and also to determine how many items would be coded for subsequent identification. At first, we thought

it would be useful to tag verbs and parts of speech. We had the problem of homographs as well as certain typing procedures necessary for optical scanning.

After considerable time and effort, we completed an eight page protocol, which, in our opinion, gave us a maximum of options and required a minimum of extra effort in typing the final hard copy. For example, we selected the apostrophe for the acute accent mark since the apostrophe is fairly similar in appearance. The acute accent mark is also the most common and for this reason we chose a lower-case keyboard character. For c cedilla (Ç) we used the comma, positioned after the C. All symbols for diacritics, in spite of practice to the contrary, were placed after the letter involved so that in the print-out, words having them would be fairly close to where they would normally fall in alphabetical order.

Typing for optical scanning.

Probably the most difficult, most expensive, and time-consuming part of the project was the data preparation of the corpus for computer input. A keypunch card operation would have been prohibitive in cost and it also presented obstacles difficult to surmount. On investigation we found that we could type for optical scanning without too much difficulty. An IBM Selectric typewriter (10 pitch) and the 12L2/12F2 typing element were all that was necessary. Data preparation for scanning is considered to be about 30 per cent faster than keypunching. Scanning, depending on the amount of material, takes about six to ten seconds a page and the cost, on a contract basis, is roughly forty dollars an hour. The average typist is capable of performing about 10,000 strokes an hour while a good operator can keypunch from 7,500 to 8,000 characters for the same period of time.

Our task was to produce a typewritten manuscript for optical character reading with as few errors as possible. We did not have a correction program easily available since we were located at some distance from the processing center. After making trials and test samples of computer print-outs to correct program errors, we attempted to proceed with typing the manuscript. On almost every page we would have one or two errors which required retyping the entire page. This was extremely discouraging since, when we corrected one error, we would often introduce others so that new and complete proofreadings were necessary. The real breakthrough came when the Office of Education granted us permission to rent an IBM Magnetic Tape Selectric Typewriter (MT/ST). Now we could produce a magnetic tape as we typed, correcting as we went, by backspacing. Upon completion of the proofreadings of the semifinal copies, the errors were corrected by going to the reference codes on the magnetic tape and then skipping lines and characters to arrive at the points where the corrections were to be made.

With the use of the MT/ST, our production increased considerably as did our accuracy. In our final computer print-out of our spoken corpus, we had about 1,000 errors in the 400,000 words typed, or roughly about a quarter of one percent. This rate by any calculation is extremely good. There were about 140 errors stemming from the process of optical scanning, coming either from faulty typewriter adjustment or the inherent characteristics of the scanner. The most common errors were those of producing a Q instead of an O, a K instead of an R, and an N instead of an M. On four occasions the scanner jumped the last four or five lines of a paragraph and once omitted almost an entire page.

This experience with our spoken corpus gave us enough confidence to attempt a similar processing of our literary manuscript. This experience were extremely frustrating. We suffered all types of problems - the blobs intended for spaces had to be blotted out by hand and the number of errors was incrediably large. Also, the errors in the text titles caused considerable problems. Correcting the errors in the literary corpus was a slow and painfull process. In the section under recommendations we are advocating a new, improved manner of computer input of linguistic data.

For our spoken corpus there were some seventy errors in informant numbers. This was a fairly serious situation since much of the value of informant numbers revolves around their being used to establish the range of how many different speakers used each individual word of the corpus. Each informant error was multiplied by the number of times that each and every word was used in the passage. On the average there were about eight informant numbers on each page, resulting in an approximate total of 7,300 errors alone from this source. Also, with the print-out in front of us, it was possible to catch many of the things we had permitted to slip by unnoticed.

Instead of using the large computer print-out to search for errors, we produced duplicate Xerox copies of the 15,000 page index for both corpora. Each corpora is now bound in 37 hard back pressure binders, which, with the five volumes of original typed transcriptions, are easily stored and readily available for consultation.

FINDINGS AND ANALYSIS

1. Description of materials produced.

As in the case of any new computer program, it was necessary to have three test runs on a small portion of the material to be processed. The final test run served as a data base for a computerized syntax analysis and is found in Part II of this report, "Toward a Computerized Syntactical Analysis of Portuguese." In this way we were able to spot

oversights and change certain features in our concordance program. With almost no facilities available to begin the project, we were most fortunate in being able to have the optical scanning and computer processing done for us by an agency of the Department of Defense.

The following were completed:

1. Key-Word-in-Context computer program (in BASIC), 120 characters, with frequency count, upper and lower case, accent marks, and informant numbers on the right margin. See appendix.

2. Frequency lists in descending order of occurrence as well as alphabetical order for both corpora. Computer programs for both.

3. Typewritten manuscripts and input tapes for both corpora.

4. Computer output tapes of the KWIC of each corpora.

5. Xerox copies of the KWIC index of each corpora (15,000 pages each), on deposit at the U.S. Naval Academy and at the School of Languages and Linguistics of Georgetown University.

6. Report by Dr. Cléa Rameh, co-investigator, entitled, "Toward a Computerized Syntactical Analysis of Portuguese," available through University Microfilms, Ann Arbor, Michigan.

7. Computer program for reverse concordance, input verification program, and a tape cartridge reading program for computer inputing from the IBM magnetic tape typewriter.

8. Translation table for alphabetizing all characters from the IBM 2741 typewriter computer terminal. This program will sort for all languages.

Our prelininary findings were based on evaluations from the KWIC index which is indispensable for any scientific analysis of the language. With all the examples of the word is question coming in the center of the page in its natural surrounding, this KWIC index provides a very rapid and efficient manner in which authentic examples of the use of the target language can be readily found for investigation. (See examples of the KWIC index in the Appendix.)

In the 400,000 word spoken language corpus, after deducting hesitations, false starts, pauses, repetitions, and errors in syntax, about 12,000 <u>different</u> words were found. In contrast, our literary corpus, which was about the same size, had 29,375 <u>different</u> words. Thus, our spoken corpus had a vocabulary range of only about forty

percent of that of the literary corpus. This figure is probably the same for other languages as well.

a. Nouns.

Our spoken corpus had some 1,685 names of persons, nicknames, and diminutives of them. About 285 geographical names were present - names of countries, states, cities, streets and so forth. Proper names, that is, commercial trade names, types, et cetera had 231 occurrences. There were at least 44 standard abbreviations (siglas) and, in spite of the fact that the conversations were between Brazil and the United States, only 146 words in English appeared, most of them in wide use in Brazil for some time. Examples are: video tape, slides, Batman, drinks, long plays, tapes, time (for team), slacks and the diminutive slaquezinho, suéter for sweater, and finally striptease. Of these 12,000 different words, a total of about 5,400, or forty-five percent, were used only once. It should be noted that all verbal forms were counted and included in the total. Shortened colloquial forms, even though they do not exist as "accepted" words, were included as were interjections. Of the 12,000 word total only 3,851, or thirty-two percent, were used five times or more.

Brazilian proper names are included, and in this study we find a predominance of double first names. There were 38 different combinations with Maria, such as Maria Cecília and Maria Helena. The double form also occurred with masculine names, but not to the same extent. There is also the use of the definite article when using proper names, except in the case of direct address.

b. Others.

While it is not surprising that the function word que was the most common (de in the literary corpus), we did not expect to find eu as number two, especially since the first person of the verb does not require a subject pronoun. As yet we have not determined whether more men than women use eu in the daily conversations. In asking a large number of Brazilians which words they regarded as most common only one was successful in guessing que and no one could imagine that eu was the second most common in Brazilian speech. Other interesting points are that the masculine definite article o is more common than the feminine (9,056 compared to 6,978), but the feminine plural as occurs more frequently than os (1,387 vs. 979). The colloquial form pra is more than three times more frequent than para. The shortened forms of tá and tou are considerably more common in speech than está and estou. The word aí appeared some 2,533 times, (as contrasted to 13 for ali, 1,138 for lá, and 3,900 for aqui. In our conversations the word aí was used to indicate the location of the second person in the conversation and distance as

such, whether close or 5,000 miles away, played no part in chosing the adverb of place. Considering the high frequency of aí, we would do well to stress its importance in our teaching.

One word that is rarely found in written form is OK of which we had 477 examples from the conversations but not a single one in the written corpus. In speech essa (459 x 67) is much more common than esta. The use of the subject pronoun ele and ela as direct object pronouns occurred 81 times.

The use of prepositions is easily found by consulting the KWIC index. While a few examples of chegar a Nova Iorque were found, there were about ten times as many occurrences of chegar em Nova Iorque. From this same source it would be fairly easy to obtain hundreds of examples of the uses of pôr and para.

Probably sensing the need for a less ambiguous possessive adjective, many Brazilian used teu and tua instead of seu and sua with você as the subject pronoun. Example: Você com a tua serenidade, teu equilíbrio - However, there were 802 seu; 971 sua, 82 teu, and 135 tua occurrences.

c. Definite articles.

From our KWIC index, we find that our informants used the definite article when talking about persons not present. Example: Eu posso falar com a Maria Lúcia? In ninety percent of the occasions the definite article was used. The same took place with prepositions to form contractions with the articles: Agora mesmo chegamos do casamento da Marieta.

The use of the definite article together with the possessive adjective was also widespread. With 1,497 occurrences of seu and sua, the definite article appeared before the possessive adjective 867 times or 57 percent of the total usage. In this group 334 examples, or 22.3 percent, combined the definite article with the preceeding preposition to form a contraction. There were 83 forms of do seu, but only 18 of de seu; 79 of da sua and 34 of de sua; 22 of ao seu with 10 of a seu; no seu 28 and em seu 3; na sua 34 and em sua 6; and finally pelo seu 18 and por seu 6; pela sua 21 and por sua 4. In summary then, of the 366 possible occasions in which our informants could have used a definite article to form a contraction with a preposition, the definite article was used 285 times or about 78 percent. From this we may conclude that in the spoken language the Brazilian prefers to use the definite article with the possessive adjective. In addition we found that seu was used immediately before names of persons some 286 times as a substitute for senhor in speech.

d.' Variants.

For the expression last week there was one example of a última semana with thirty-five for a semana passada. For next week the count was sixteen for na próxima semana, one for a semana próxima and fifty-seven for a semana que vem. To give the day of the month the common form was: no dia dois.

There were forty-five examples of o telefonema, but a telefonema appeared nine times. Questão without pronouncing the u occurred sixty-four times (being used to a much greater extent by the younger generation) while qüestão had nineteen examples. Faz favor was found only twenty-nine times compared to 165 sentences with por favor.

For the Brazilian equivalent of tonight, there were fifteen occurrences of hoje à noite, three of hoje de noite, but only one for esta noite. There were nineteen examples of the salutation um boa noite contrasting with twenty-four of uma boa noite. Roughly ten percent of the occurrences of outro were preceded by the indefinite article um and the same held true for outra with uma. Example: E depois eu dou a você uma outra resposta.

With several nouns there are the two possible choices in forming expressions as in tenho muitas saudades or estou com muitas saudades. In our corpus Brazilians prefered the latter forms by 100 to one, and in this very expression the women would usually say: estou morrendo de saudades. For being hot or cold the form was: estou sentindo muito frio. A cursory check to see the degree to which the Brazilian uses the subject pronoun eu, showed that eu was used in over half the occasions in which it could have been spoken. There was a tendency not to repeat it if it had been used in the first clause of a sentence. Certain sociologists, noting that the Brazilian male had a tendency for saying eu e a minha mulher, interpreted this as a case of machismo. However, in our corpus the women also put themselves first when speaking as in: Eu e Roberto já fizemos tudo pra ver se ela ... and also Eu e Ney morremos de saudades.

Finally, there were ninety-nine occurrences of não tem problema while the more "elegant" não há problema appeared some sixty-one times. Of course, there are many more features of the spoken language which can be checked and counted. It would be useful to have a spoken corpus of about a million words because, in some cases, there are not enough examples so that accurante findings can be determined with precision.

c. Description of Verbal Forms.

It is somewhat difficult to establish just what consitiutes a verb or verbal form. In any case the following guidelines were used for the frequency count.

1. The infinitive includes all those forms used in the conversational future, such as: Eu vou ver. No distinction is made between this group and the personal infinitives of the first and third persons singular.

2. The present participles include all forms regardless of function.

3. The past participles are divided into three groups: a) those used with estar in any tense, b) those used with ser, and c) all other examples including those forming compound tenses and those functioning adjectively other than with estar.

4. Progressive tenses are those forms used with the various tenses of estar and the present participle.

5. The conversational future, formed with the present of ir and the infinitive, is counted separately even though the forms of ir and the infinitives are also counted individually. No distinction is made between vamos "let's" and vamos "we will."

6. The classification imperative substitute designates those forms of the third person singular of the present indicative that are popularly used for commands with você as the subject pronoun.

7. The emphatic future is the form composed of the present indicative of haver and the prepostion de and a dependent infinitive. Eu hei de saber.

Totals.

There were a total of 89,209 different items, but it should be stated that compound forms were usually counted three times, one for each part and another for the complete unit. Deducting these duplicates produced a total of 81,091 verbal forms. The table on the following pages gives a breakdown according to the various verbal forms.

VERBAL FORM		OCCURRENCES	PERCENT	ADJUSTED*
Infinitive		15,385	17.25	18.97
Personal Infinitive		215	0.24	0.27
Conversational Future		3,695	4.14	
Emphatic Future		13	0.01	
Present Participle		4,613	5.17	5.69
Past Participle		3,371	3.78	4.16
With estar	643			
ser	515			
others	2,213			
Present Indicative		32,062	35.94	39.54
Preterite		12,309	13.80	15.18
Imperfect		2,266	2.54	2.79
Future Indicative		1,587	1.78	1.96
Conditional		742	0.83	0.92
Present Subjunctive		1,612	1.81	1.99
Imperfect Subjunctive		579	0.65	0.71
Future Subjunctive		1,559	1.75	1.92
Infinitive Perfect		180	0.20	
Present Perfect Indicative		604	0.68	
Past Perfect Indicative				
With ter		139	0.15	
haver		48	0.05	
Simple Plu-perfect		1	0.00	0.00
Future Perfect Indicative		9	0.01	
Conditional Perfect		12	0.01	
Present Perfect Subjunctive		93	0.10	
Past Perfect Subjunctive		30	0.03	
Future Perfect Subjunctive		7	0.01	

VERBAL FORM (cont.)	OCCURRENCES	PERCENT	ADJUSTED*
Imperative Substitute (colq)	2,875	3.22	3.55
Command - 3rd person	1,729	1.94	2.13
Command - tu (2nd person)	134	0.15	0.17
Infinitive Progressive	50	0.06	
Present Progressive	2,779	3.12	
Preterite Progressive	16	0.02	
Imperfect Progressive	165	0.18	
Future Progressive	36	0.04	
Conditional Progressive	3	0.00	
Present Subjunctive Progressive	39	0.04	
Imperfect Subjunctive Progressive	6	0.01	
Colloquial forms - ·			
esteje, estejem	8	0.01	
seje	7	0.01	
vim (substitutes for vir)	37	0.04	

* Based on 81,091 verbal forms, compound forms were only counted once.

As can be seen from the distribution of verbal forms, the present indicative is by far the most common tense in spoken Brazilian Portuguese. Counting all the present indicative verbal forms, including those forming perfect and progressive tenses, we find some 32,000 examples or a little less than thirty-six percent of the total. Subtracting the compound forms increases the present indicative tense up to nearly forty percent. Below are some of the verbs which show a preponderant usage in the present indicative.

VERB	OCCURRENCES	NO. IN PRESENT	PERCENT
estar	9,474	7,370	77.8
ir	7,487	5,582	74.8
ser	7,182	4,806	66.9
ter	4,099	2,292	55.5
querer	2,602	1,509	58.0
saber	2,451	1,227	50.1
poder	2,010	1,001	49.8
dever	819	663	80.9

Next in importance came the infinitive with almost 19% of the occurrences. The conversational future (3,695 ex.) was more than three times as common as the future indicative (1,587 ex.). Also, some verbs such as querer were never used in the future, nor in the conditional, because, according to many Brazilians querried, of the rather harsh sounds created in pronouncing these verbal forms.

There were 12,309 preterite forms, but 1,029 of them were viu, a common corruption which is used as an interjection seeking confirmation of a previous statement or perhaps the equivalent of "Y' hear?" Chegar, dizer, entender, escrever, escutar, fazer, falar, mandar, ouvir, pedir, receber, responder, seguir, sofrer, and telefonar are verbs which have, by far, more of their finite forms in the preterite than in any other tense.

In the spoken language the imperfect is much less frequent than in the written form. Of the 2,266 imperfects some 1,939 of them are confined to only eight verbs:

querer	591	ir	174
estar	307	ser	172
ter	300	haver	67
(shortened form of estar)		poder	63
'tava	203	saber	62

It is evident that the high frequency of querer in the imperfect stems, in part, from its use as a conditional substitute. There were only 26 examples of the preterite quis. In contrast, poatar has almost no imperfect forms, but it was the most common verb in the condi onal tense.

For the great majority of verbs there were very few or no occurrences in the future indicative. Estar (161), ir (151), ser (137), and poder (104) were the most common.

Only a small number of verbs had conditional forms:

gostaria	231	iria	26
poderia	73	ficaria	17
teria	38	deveria	14
pediria	28	estaria	12
		viria	11

In contrast to English and to Spanish,[14] spoken Brazilian Portuguese makes relatively little use of the present perfect tense. The occurrences were confined to a few verbs, the most common being:

escrever	92	fazer	30
receber	67	chegar	23
ser	61	ir	17
ter	48	mandar	15
estar	30	sair	11

The remaining perfect (compound) tenses are even less frequent in the spoken language. To form the past perfect indicative the auxiliary ter was used 139 times compared to 48 uses of a form of haver. However, our literary corpus showed an almost equal usage of the forms of ter and haver. Only one example was found of the simple (or literary) pluperfect tense.

All the subjunctive tenses together came to about five percent of the total number of occurrences. Some of the more common verbs used in the present subjunctive were:

ter	185	dar	47
ser	113	chegar	46
estar	104	vir	44
poder	101	dizer	38
ir	77	escrever	33

The occurrences of the future subjunctive were confined to a rather limited number of verbs. Of the 262 examples of quiser, some 125 came from the expression: Se Deus quiser. Most of the remainder consisted of some form of: Se quiser.

se, quando, assim que, logo que (which also introduces the present subjunctive), o que (or whatsoever), o mais que, qualquer coisa que, o primeiro que, and sempre que were used to introduce the future subjunctive. Most of the future subjunctives were found in the following verbs:

querer	262	ir	55
poder	190	precisar	51
ter	136	vir	49
ser	129	haver	45
chegar	110	conseguir	31
estar	101	achar	29

The colloquial command form, which we call the imperative substitute, is actually the third person singular of the present indicative. It is half again as common (2,875 vs. 1,729) as the standard third person command for você, o senhor, and a senhora. It should be noted, however, that 1,027 of these occurrences were for the attention getter olha. The most common imperative substitutes were:

olhar	1027	esperar	75 ('pera 44)
dizer	208	fazer	56
dar	202	deixar	48
falar	196	ficar	45
ver	105	telefonar	39
mandar	99	responder	38
escutar	90	pedir	38

A look at some verb forms in the KWIC index brings out some interesting things. There are 66 examples of buscar, but not one of them is a finite verb - much in contrast to the near complete set of verbal forms for procurar. A typical example would be: Eles vão te buscar no hotel.

Then there is the question of tenho de falar or tenho que falar. Our findings were very conclusive - tenho que was used almost 99 percent of the time in the spoken corpus. A check of our literary corpus, which is of similar size, produced an equal number of examples for both expressions.

Unusual irregular forms appeared: seje, esteje, and estejem as well as their shortened variants of teje and tejem. This is unconcious overcorrection. Example: Desejo que tudo teje correndo bem. There were also three examples of escrevido and 18 of deixa eu. Subjects and verbs failed to agree in number in many sentences. Then there were 37 cases in which vim was used to replace vir as the infinitive in examples such as: Você tem que vim em dezembro de qualquer maneira.

Finally, a word regarding the person of the verb. From the KWIC index we find that the third person singular is the most common, closely followed by the first person singular. The plural forms, however, have only about one-tenth the frequency of the singular forms.

In our spoken corpus você was the common form of address (6,771 occurrences and also used 132 times as a direct object pronoun) in contrast to o senhor (419 times) and a senhora (742 times). Son and daughters usually used você when addressing a parent, but would use a senhora when speaking to their mother-in-law. Most of the individuals who used o senhor and a senhora when addressing their own parents individually, used vocês when speaking or referring to both of them. In the plural this contrast was even greater - we had 1,479 occurrences of vocês but only twenty-one of os senhores. Also, there were 153 examples of tu, but not a single vós. Once again, those who used tu in the singular also used vocês in the plural.

On the following pages are the lists by descending order of frequency and also by alphabetical order for all verbs having forty or more occurrences. On the alphabetical list there is a notation as to the most common verb form found for each different verb. Complete lists of verb-form frequencies of all 650 verbs are available to scholars on request.

Conclusions and Recommendations.

The findings presented in this report are only some of the highlights taken from the word index of the spoken corpus and the frequency lists. A syntactical analysis remains to be done. Now that a spoken corpus and also a literary corpus are available in concordance form, it is hoped that others will take advantage of this opportunity. For the present the KWIC indexes will be available at the U.S. Naval Academy and also at the School of Languages and Linguistics of Georgetown University. We are attempting to produce microfiche copies as well as computer tapes of the KWIC lists.

During the period of the contract several events took place. Methods for inputing manuscripts to the computer were developed, time-sharing systems became available, and new hardware could effectively deal with the large volume of data to be processed. In other words, were the project started over today, there would be many changes in methods and many of the difficulties we experienced would no longer be present.

Based on our experience optical scanning should be rejected as a means for inputing data to the computer. The error rate is too high and the correction procedure is too cumbersome. A key punch operation is also not the most efficient method. Instead, one of the best ways to input data is by a time-sharing system from a remote terminal. The IBM 2741 teletypewriter terminal is one of the most versatile since it has the capability of having the interchangeable sphere, a feature which will work for various different languages. The 2741 then works on a disk pack and corrections can be made from the terminal. Also, a computer verification program can be used to catch some of the more obvious errors. But perhaps the best method is to have the manuscripts typed twice, once each by two different persons. A comparison programs reads both files, printing all discrepancies. The human proofreader then has no more to do than to select the correct form. Of course, the correct material is not touched and additions or deletions can be made without disturbing the original data. Once the data bank is on the disk pack a very efficient edit software package provides for making text replacements, changing thousands of examples of a unique string by one single command. Such a feature is most important for during the period of our contract the Brazilian government changed the official orthography of Portuguese. Once we get the magnetic tapes on our system, we should be able to make the changes without too much trouble.

In preparing the manuscript for computer input there should be the least amount of coding possible. Coding for syntactical items only leads producing more errors and additional programming problems. Since the items fall out together in the KWIC index and also in the reverse alphabetical concordance, very little is gained. The possible exception is the case of homonyms as, for example, the preterite forms of ser and ir and the preposition a. Here a non-print character can be affixed to the word in question and, while it will not appear, all occurrences will be sorted separately.

Although using upper and lower case adds considerably to computer running time, this feature will provide for a distinction between proper and common nouns. At the same time there is an exact reproduction of the original text. The use of accent marks adds to the problem since a special program must place of words with an accent mark immediately after a similar one without the accent mark as in dictionary form. Without a special compensating computer program words with accent marks will be alphabetized out of order and will be the first words after the beginning of each letter.

ENDNOTES

[1] H.A. Gleason, An Introduction to Descriptive Linguistics (Revised edition, New York, Holt, Rinehart and Winston, 1967), p. 10

[2] Ibid., p. 11.

[3] Earl W. Thomas, The Syntax of Spoken Brazilian Portuguese (Nashville, Vanderbilt Univ. Press, 1969).

[4] Radio conversation with Prof. Henry Hoge, Rio de Janeiro, May 22, 1968.

[5] Charles F. Hockett, A Course in Modern Linguistics (New York, Macmillan, 1967), p. 546.

[6] Henry Kucera and W. Nelson Francis, Computational Analysis of Present-day American English (Providence, R.I., Brown Univ. Press, 1967), p. 371.

[7] Charles R. Brown, Wesley M. Carr, and Milton L. Shane, A Graded Word Book of Brazilian Portuguese (New York, F.S. Crofts & Co., 1945).

[8] Charles R. Brown and Milton L. Shane, Brazilian Portuguese Idiom List: Selected on the Basis of Range and Frequency of Occurrence (Nashville, Vanderbilt Univ. Press, 1951), p. 2.

[9] Ibid., p. 6.

[10] John R. Kelly, "A Computational Frequency and Range List of Five Hundred Brazilian-Portuguese Words," Luso-Brazilian Review Vol. VII, No. 2 (Dec. 1970), pp. 104-113.

[11] John Duncan, "Frequency Dictionary of Portuguese Words," Diss, Stanford Univ., 1971.

[12] Leslie Kish, Survey Sampling (New York, Wiley, 1965).

[13] Thomas, p. xii.

[14] William E. Bull et al., "Modern Spanish Verb-Form Frequencies," Hispania, Nov. 1947, p. 458.

VOCÁBULO

1	estar	9474	51	correr	231
2	ir	7407	52	continuar	220
3	ser	7182	53	chamar	217
4	ter	4093	54	providenciar	203
5	falar	3295	55	entrar	193
6	mandar	2782	56	levar	192
7	querer	2602	57	trazer	190
8	saber	2451	58	encontrar	185
9	dizer	2186	59	aproveitar	182
10	poder	2010	60	sentir	169
11	passar	1921	61	começar	166
12	dar	1817	62	tentar	157
13	fazer	1802	63	parecer	156
14	ficar	1772	64	trabalhar	154
15	receber	1706	65	pensar	151
16	chegar	1392	66	viajar	150
17	escrever	1374	67	explicar	179
18	ouvir	1260	68	conversar	173
19	entender	1176	69	esquecer	172
20	vir	1169	70	despedir	167
21	olhar	1144	71	pagar	153
22	esperar	1067	72	entregar	151
23	ver	1039	73	andar	146
24	viu ?	1029	74	confirmar	141
25	dever	819	75	repetir	140
26	haver	746	76	comunicar	139
27	pedir	737	77	tratar	139
28	sair	684	78	acontecer	135
29	achar	671	79	tirar	136
30	precisar	617	80	compreender	134
31	telefonar	583	81	acabar	129
32	voltar	491	82	estudar	125
33	preocupar	473	83	enviar	123
34	levar	469	84	botar	118
35	gostar	445	85	desejar	115
36	conseguir	413	86	mudar	112
37	avisar	411	87	depender	110
38	resolver	407	88	pegar	110
39	deixar	371	89	arranjar	108
40	escutar	336	90	informar	106
41	aguardar	333	91	embarcar	104
42	responder	331	92	faltar	103
43	comprar	325	93	preparar	101
44	procurar	320	94	terminar	100
45	tomar	290	95	marcar	99
46	agradecer	246	96	combinar	96
47	morrer	233	97	transmitir	96
48	seguir	229	98	demorar	95
49	perguntar	217	99	lembrar	95
50	contar	215	100	conhecer	90

VERBOS (continued)

101	destinar	99
102	melhorar	98
103	arrumar	95
104	por	95
105	ajudar	84
106	pretender	83
107	perder	81
108	adorar	80
109	passear	76
110	herdeitar	74
111	buscar	73
112	remeter	72
113	chorar	71
114	jantar	71
115	morar	67
116	atender	64
117	apanhar	61
118	ocupar	61
119	colocar	60
120	descansar	60
121	ditar	59
122	adiantar	56
123	crer	52
124	dormir	51
125	comer	50
126	cuidar	50
127	parar	50
128	recuperar	50
129	ler	49
130	custar	48
131	interessar	48
132	gastar	44
133	abusar	43
134	alugar	43
135	acertar	42
136	aprender	41
137	despachar	40

VERBOS:

abusar	53	inf 17
acabar	129	pret 63
acertar	42	pres ind 13
achar	671	pres ind 495 (acho 398)
acontecer	136	pret 45
acreditar	74	pres ind 69 (all are acredito)
adiantar	50	pres ind 17 (all are adianto)
adorar	80	pret 34
aparecer	246	pres ind 96
aguardar	383	pres part 109
ajudar	84	inf 35
alugar	43	inf 12, pret 11
andar	146	pres ind 42
apanhar	61	inf 38
aprender	41	inf 12
aproveitar	192	inf 49
arranjar	108	inf 32
arrumar	85	inf 23
atender	62	inf 22
avisar	411	inf 159, imp sub 50
bater	116	inf 43, pret 36
buscar	73	inf 57, used only in inf & conv. fut.
chamar	307	inf 52, pres ind 30
chegar	1312	pret 354, inf 239, pres ind 106
chorar	71	pres part 20, inf 17
colocar	60	inf 24, pret 11
combinar	96	past part 30
começar	148	pret 47
corar	50	pres part 17
comprar	325	inf 132, pret 80
compreender	134	pret 63, pres ind 34
comunicar	139	inf 64
confirmar	141	inf 42
conhecer	90	pres ind 35
conseguir	416	pret 159
contar	215	pres part 80
continuar	210	pres ind 75
conversar	171	inf 66
correr	211	pres proc 76
crer	52	creio only form
cuidar	50	commands 16
custar	48	pres ind 14
dar	1017	inf 552, pret 202, imp sub. 202
deixar	371	inf 129, pret 72
demorar	95	inf 34
depender	110	pres ind 42 (depende 36)
descansar	60	past part 43
desejar	115	pres ind 70
desligar	90	inf 42
despachar	40	inf 12

VERBFORM (continued)

despedir	147	inf 74 (como fut 48)
dever	819	pres ind 663 (deve 467)
dizer	33	past part 25
dizer	2166	inf 865, pres 403, com 356 (diga 294)
dormir	51	inf 19
encarar	156	inf 31, pres ind 30
encontrar	156	inf 54, pret 42
entender	1176	pret 566 (entendi 413, entendeu 236)
		past part 291
entrar	189	inf 77
entregar	181	inf 50, pret 37
enviar	123	inf 26
escrever	1374	pret 405, in. 304
estudar	356	pret 184, escuta imper. sub 80
esperar	1067	pres inf 275, past part 240 (*para 44)
esquecer	172	commands 61, pret 38
estar	9474	pres ind 7370 (tn 5194, tou 660)
estudar	125	pres part 44
explicar	179	pres part 45
falar	3359	inf 1350
faltar	103	pres ind 44 (only 3rd persons)
fazer	1802	inf 611, pret 233
ficar	1772	inf 565
gastar	44	inf 19
gostar	445	cond 239 (gostaria 231), pret 104
haver	745	only 3rd sing., ha 435, houve 95
		houver 45
informar	106	inf 30
interessar	48	past part 19
ir	7407	pres 5562
jantar	71	inf 33
lembrar	95	pres ind 36
ler	49	pret 18
levar	469	inf 168, pret 77
ligar	198	inf 168, pret 77
mandar	2782	pret 724, inf 628
marcar	99	inf 22
melhorar	86	pret 20
morar	87	pres ind 25
morrer	233	pres part 100
mudar	112	pret 40, inf 36
ocupar	61	inf 14
olhar	1144	olha (attention getter) 1027, olhe 87
ouvir	1260	pret 633
pagar	153	inf 70
parar	50	inf 13
parecer	186	pres ind 163 (parece 161)
passar	1921	pres prog 305, inf 299
passear	76	inf 22
pedir	737	pret 316, pres ind 143

Vocabulário (continuação)

pegar	110	inf 36
pensar	181	pres part 46
perder	81	pret 24
perguntar	217	inf 52
poder	2010	pres ind 1001 (pode 603)
ter	65	pret 22
precisar	617	pres ind 346 (preciso 210, precisa 136)
preocupar	473	commands 196, past part 134
preparar	101	pres part 33
prender	438	pres ind 51
procurar	320	inf 118
providenciar	203	inf 68
querer	2602	pres ind 1509, imper 591, fut sub 262
receber	1706	pret 1093 (recebi 548, recebeu 356)
recuperar	50	pres part 17
roubar	72	pret 19, inf 19
repetir	140	inf 52
resolver	407	inf 127
responder	331	inf 81, pret 60
saber	2451	pres ind 1227 (sei 697), inf 1005
sair	604	inf 201
seguir	229	pret 67
sentir	139	pres ind 55
ser	7182	pres ind 4806 (é 4210, são 208)
telefonar	503	inf 178, pret 124
tentar	187	inf 48 (conv fut 39), pres 37
ter	4099	pres ind 2292 (tem (sing.) 1340)
terminar	100	inf 34
tirar	136	inf 61
tomar	259	inf 109
trabalhar	184	inf 42, pres prog 39
transmitir	96	inf 33
tratar	139	inf 31, pres ind 27
trazer	193	inf 93, pret 36
ver	1039	inf 527, (vou ver 66, vamos ver 48)
viajar	180	inf 78
vir	1169	pres ind 481, vim for vir 37
viu ?	1029	interjection (form of ver)
voltar	491	inf 171, pres 74

DISTRIBUTION OF INFORMANTS BY PROFESSION

Code	Profession	Male	Female	Total
A	Aviation	27	1	28
B	Banking	10	2	12
C	Businessman	15	1	16
D	Diplomat	10	2	12
E	Army	21	0	21
F	Clerk (office)	9	25	34
G	High government official	7	0	7
H	Housewife*	0	335	335
I	Engineer (professional)	14	0	14
J	Journalist (newspaper)	2	0	2
K	Child	6	9	15
L	Lawyer	6	0	6
M	Physician	19	1	20
N	Navy	36	0	36
O	Writer	0	1	1
P	Schoolteacher	--	13	13
Q	Security police		0	4
R	Radio & TV personnel	12	1	13
S	Student	43	72	115
T	Clergy	1	0	1
U	University professor	12	1	13
V	Farming	3	0	3
W	Native of Portugal	1	1	2
X	Not identified	97	0*	97
Y	Business executive	9	0	9
Z	Non-natives	3	5	8
	Total	367	470	837

* Many of the women who could not be identified as to profession were placed in the housewife category.

Author and Selections List

	Code	Pages	Selection
1.	AT	1a-21	Adolfo Tilio, Bahia, 1919:
	AW	50-62	Contos vivos. Rio de Janeiro,
	AW	97-108	Ed. Civilização Brasileira, 1962. 156 pp.
2.	CA	7-29	Andrade, Carlos Drummond de,
	CA	58-82	Minas, 1902: Contos de aprendiz.
	CA	125-143	3a Ed. Rio de Janeiro, Editora do Autor, 1963. 207 pp. (First Edition, Rio, José Olympio, 1951)
3.	RB	37-67	Braga, Rubem, Espírito Santo,
	RB	99-123	1913: Ai de ti, Copacabana. 4a
	RB	191-211	Ed. Rio de Janeiro, Editora do Autor, 1960. 222 pp. (1a Ed. 1960).
4.	JC	1-23	José Condé, Pernambuco, 1916:
	JC	53-73	Um ramo para Luiza. 3a Ed. Rio,
	JC	114-136	Editora Civilização Brasileira, 1961. 145 pp. (1a Ed.: Rio de Janeiro, Civilização Brasileira, 1959).
5.	CC	17-36	Cony, Carlos Heitor, Rio, 1926:
	CC	86-99	Antes o verão. Rio de Janeiro,
	CC	127-139	Editora Civilização Brasileira, 1964. 171 pp.
6.	AD	1-13	Dourado, Waldemiro Autran, Minas,
	AD	38-48	1926: Uma vida em segredo. Rio de
	AD	84-98	Janeiro, Editora Civilização Brasileira, 1964. 103 pp.
7.	OF	3-15	Faria, Octávio de, Rio, 1908:
	OF	107-119	Angela ou as Areias do Mundo
	OF	261-272	(O anjo de pedra, IV). Rio de Janeiro, José Olympio, 1963. 390 pp. (Tragédia Burguesa, IV)

Cód.	Pagsº	Referência
8. [illegible]	[illegible]-10	Nogueirede, Guilherme, São Paulo,
[illegible]	[illegible]-41	1914: O [illegible] do [illegible]. Rio,
[illegible]	60-63	Editora Civilização Brasileira,
[illegible]	94-109	1964. 187 pp.
[illegible]	130-145	
9. [illegible]	17-24	Fonseca, Rui Antônio Carvalho da:
[illegible]	[illegible]-82	Sete milênios, Rio de Janeiro,
[illegible]	103-115	Livraria Freitas Bastos, 1961.
[illegible]	135-157	219 pp.
[illegible]	191-204	
10. LM	(23 crônicas)	Henriques, Luís, Barra-Pernambuco,
		1926: Menino entrega telegrama.
		Rio, Pongo Brasileiro, 1964.169 pp.
11. [illegible]	1-10	Ivo, Lêdo, Alagoas, 1914: O
[illegible]	29-43	sobrinho do general. Rio de
LI	76-91	Janeiro, Ed. Civilização Brasi
		leira, 1964. 124 pp.
12. AL	31-45	Leite, Ascendino, Paraíba, 1915:
AL	89-103	A prisão. Rio de Janeiro, Ed. O
AL	130-145	Cruzeiro, 1960. 213 pp.
13. OL	1-14	Lins, Osman, Pernambuco-São Paulo,
OL	94-109	1924: O [illegible] de primeira
OL	151-165	viagem. Rio de Janeiro, Editora
		Civilização Brasileira, 1963.
		165 pp.
14. [illegible]	39-50	Lopes, Moacir, Ceará [illegible]:
[illegible]	97-108	A [illegible] da cada [illegible]. 2ª Ed. Rio
[illegible]	225-239	de Janeiro, Editora Civilização
		Brasileira, 1962. 239 pp. (1ª Ed.
		Rio, 1959)
15. JM	11-21	Martins, João, Bahia, 1913:
JM	112-114	Os [illegible]. Rio de Janeiro,
JM	[illegible]-137	Edições O Cruzeiro, 1964, 242 pp.
JM	140-159	
JM	176-188	

	Code	Pages	Selection
16.	MO	15-28	Montello, Josué, Maranhão, 1917:
	MO	72-85	O labirinto de espelhos. 2a Ed.
	MO	126-137	São Paulo, Livraria Martins, 1962. 161 pp. (1a Ed. Rio, José Olympio, 1952).
17.	SM	13-25	Moraes, Santos (José), Bahia,
	SM	95-108	1920: Os filhos do asfalto.
	SM	174-186	Rio de Janeiro, Ed. José Alvaro, 1964.
18.	EN	21-31	Nascimento, Esdras do, Piauí,
	EN	56-67	1934: Solidão em família. Rio
	EN	99-110	de Janeiro, Editôra Civilização
	EN	132-145	Brasileira, 1963. 233 pp.
19.	SP	3-17	Paezzo, Sylvan: Época dos tristes.
	SP	25-38	Rio de Janeiro, Editôra Civili
	SP	54-66	zação Brasileira, 1964. 121 pp.
	SP	79-92	
	SP	99-111	
20.	PP	37-60	Porto, Sérgio (Preta, Stanislaw
	PP	76-97	Ponte), São Paulo: Primo Alta-
	PP	150-171	mirando e elas. 2a. Ed. Rio de Janeiro, Editôra do Autor, 1962 (1st. Ed. 1962). 206 pp.
21.	RR	19-31	Ramos, Ricardo, Alagôas, 1929:
	RR	117-128	Os desertos. São Paulo, Edições Melhoramentos, 1961. 168 pp.
22.	OR	5-17	Resende, Otto Lara, Minas, 1922:
	OR	78-90	O braço direito. Rio de Janeiro,
	OR	185-197	Editôra do Autor, 1963. 233 pp.

Code	Pages	Selection
23. NR	19-24	Rodrigues, Nelson, Pernambuco, 1912: 100 contos escolhidos: a vida como ela é. Vol. I, Rio de Janeiro, J. Ozon, 1961. 316 pp.
NR	25-30	
NR	37-42	
NR	49-54	
NR	67-72	
NR	73-78	
NR	137-142	
NR	173-178	
NR	185-190	
NR	197-202	
24. FS	9-22	Sabino, Fernando, Minas, 1923: O encontro marcado. 5a Ed. Rio de Janeiro, Ed. Civilização Brasileira, 1960. 287 pp.
FS	49-64	
FS	93-106	
FS	151-164	
25. DT	5-20	Trevisan, Dalton, Paraná, 1926: Morte na praça (contos). Rio de Janeiro, Editôra do Autor, 1964. 115 pp.
DT	35-50	
DT	79-92	
26. JV	9-18	Vasconcelos, José Mauro de, Rio Grande do Norte: Doidão, São Paulo, Exposição do Livro, n.d. (1963) 102 pp.
JV	39-53	
JV	73-86	
27. EV	24-36	Veríssimo, Érico, R.G. do Sul, 1905: O arquipélago. Vol. (O Tempo e o Vento, 3a Parte). Pôrto Alegre, Editôra Globo, 1961. 304 pp.
EV	124-137	
EV	248-260	

IDENTIFICATION TAGS AND CODES

1. Verbs.

a. The preterite forms of the verb SER will be typed with a dart ∇ following. The same applies to all forms derived from the third person plural of the preterite of SER.

Foi bom eu ter falado com você.
FOI∇/BOM/EU/TER/FALADO/COM/VOCE=/./

A bagagem ainda não foi despachada.
A/BAGAGEM/AINDA/NA+O/FOI∇/DESPACHADA/./

que fôsse retirado até segunda ordem
QUE/FO=SSE∇/RETIRADO/ATE'/SEGUNDA/ORDEM/

Se fôr possível Dona Odete ...
SE/FO=R∇/POSSI'VEL/DONA/ODETE/

Os resultados dos exames foram ótimos.
OS/RESULTADOS/DOS/EXAMES/FORAM∇/O'TIMOS/./

b. The following forms of VER will be followed by the ∇ dart symbol: via; the preterite forms viu, viste, vimos and viram and the forms derived from viram.

Êle viu o amigo.	E=LE/VIU∇/O/AMIGO/./
Vimos ela no jardim.	VIMOS∇/ELA∇/NO/JARDIM/./
Se você vir Maria ...	SE/VOCE=/VIR∇/MARIA/./

c. A verb form with an attached pronoun will be typed without a hyphen.

cuide-se	CUIDE/SE∇	mudou-se	MUDOU/SE∇/
alistar-se	ALISTAR/SE∇	chama-se	CHAMA/SE∇/

Exception: Shortened infinitives.

levá-lo	LEVA'-/LO/	recebê-lo	RECEBE=-/LO/
ouvi-lo	OUVI-/LO/	ajudá-la	AJUDA'-/LA/

2. Articles and Pronouns.

a. Articles used as pronouns and their contractions will be identified by a ∇ directly after the pronoun, except when the article is followed by que.

o daqui	O∇/DAQUI/
o do rio	O∇/DO/RIO/
a de português	A∇/DE/PORTUGUE=S/
queres que eu o espere	QUERES/QUE/EU/O∇/ESPERE/
no de abril	NO∇/DE/ABRIL/

b. The reflexive pronouns ME, TE, SE, and NOS will be followed by the ∇ dart symbol.

eu nunca me separei dêle	EU/NUNCA/ME∇/SEPAREI/DE=LE/
eu me lembro	EU/ME∇/LEMBRO/
tenho me interessado	TENHO/ME∇/INTERESSADO/
estamos nos combinando	ESTAMOS/NOS∇/COMBINANDO/
pra se despedir	PRA/SE∇/DESPEDIR/
não se preocupe	NA+O/SE∇/PREOCUPE/

c. Colloquial direct object pronouns ÊLE, ELA, ÊLES, ELAS, VOCÊ, VOCÊS are identified with a ∇ following.

eu seguro êle mais essa ..	EU/SEGURO/E=LE∇/MAIS/ESSA/
eu despacho êle para ...	EU/DESPACHO/E=LE∇/PARA/
conhecem você	CONHECEM/VOCE=∇/
eu ouvi você bem	EU/OUVI/VOCE=∇/BEM/

3. Contractions.

a. Except as noted below, contractions are typed as normally written without special identification.

nisso	NISSO
desta	DESTA

b. The contraction NOS, to be differentiated from the object pronoun (no special identification) and the reflexive pronoun (followed by ∇), is printed with an & ampersand following.

nos cursos	NOS&/CURSOS/
nos meninos	NOS&/MENINOS/

c. For the contraction Â and ÂS, see section on accent marks.

4. Prepositions and Conjunctions.

a. The prepostion A will be distinguished from the articles or the pronoun by the & symbol.

começar a andar	COMEC,AR/A&/ANDAR/
está a caminho	ESTA'/A&/CAMINHO/
a viagem a Nova Iorque	A/VIAGEM/A&/NOVA∇IORQUE/
Eu escrevi a Rubens	EU/ESCREVI/A&/RUBENS/
a respeito	A&/RESPEITO/

b. The interrogative por que is written with one interspace, with the ∇ symbol linking each of the elements.

por que	POR∇QUE

5. Interjections.

a. Interjections that are used for a pause will be followed by the ▽ symbol. All others will be typed without special identification. The following are some that occur:

joy, admiration	AH/EH/OH/
pain	AI/UI/
aversion	IH/CHI/
plea	O/
pause	AH▽/EH▽/IH▽/OH▽/UH▽/
call	O=/
answer	OI/
O.K.	OK/

6. Multi-word expressions.

a. When an expression, such as a title, proper name or multi-unit number is composed of more than one word, the elements are linked with the ▽ symbol.

Rio de Janeiro	RIO▽DE▽JANEIRO
Maria Teresa	MARIA▽TERESA
Pôrto Alegre	PO=RTO▽ALEGRE
trezentos e cinco	TREZENTOS▽E▽CINCO
et cetera	ET▽CETERA

b. Exceptions are made for lengthy expressions (over 14 characters) and for those which are not bona fide titles.

7. Abbreviations.

a. Standardized abbreviations are typed as pronounced.

CAPES/	COMSAT/
EMFA/	A/FAB/
INTELSAT/	A/OEA/
A/ONU/	A/VARIG/

8. Accent marks.

a. The following symbols, which follow the letters, are used for accent and diacritical marks in Portuguese:

´	'	é	E'
`	#	às	A#S
^	=	êle	E=LE
~	+	não	NA+0
ç	,	canção	CANC,A+0

9. Punctuation.

a. The system of punctuation adopted in transcribing the conversations attempts to reflect more closely the actual speech patterns rather than conform to the standard rules of sentence punctuation.

b. A hyphenated word will remain hyphenated, except as shown in paragraph 1c.

sexta-feira	SEXTA-FEIRA
capitão-de-corveta	CAPITA+O-DE-CORVETA

c. The shortened form of weekdays will retain the hyphen.

a sexta	A/SEXTA-/

10. False starts, incomplete words, pauses, etc.

a. A single word repeated once or several times as the speaker ponders what to say next will be linked with the repetition, in the same interspace, by a dart ▽ .

eu eu encerro	EU▽EU/ENCERRO/

A single word repeated for emphasis will not be linked.

não senti nada nada nada	NA+O/SENTI/NADA/NADA/NADA/

b. A false start involving part of one word immediately substituted by another will be typed as follows:

preci	/(PRECI-/

c. A false start involving a complete word immediately substituted by another will be typed as follows:

eu você deve	/(EU/VOCE=/DEVE/

d. A false start that results in an incomplete thought or sentence will be typed without special identification.

eu vou eu quero saber	/EU/VOU/EU/QUERO/SABER/

e. A substantial pause, for any reason, will be shown by three suspension dots in the interspace. The symbol ▽ will precede.

f. Laughter is indicated: HA▽HA▽HA/

11. Colloquial forms.

a. A partial word reflecting colloquial speech habits rather than a false start is typed as follows:

'tendi	/EN)TENDI/
'pera	/ES)PERA/
'ta (for ESTAR)	/ES)TA(R/
'tão	/ES)TA+O/
'tava	/ES)TAVA/
'tive	/ES)TIVE/

b. However, the following colloquial forms are typed verbatim, with symbols appearing only as heretofore indicated.

cê	CE=	cês	CE=S
né	NE'	pa	PA
po	PO	pra	PRA
pras	PRAS	pro	PRO
pros	PROS	tá	TA'
tás	TA'S	tou	TOU
té	TE'	viste	VISTE
viu (ouvir)	VIU		

12. Word variants.

a. The following pronunciation variant will not be normalized but will be typed phonetically thus:

NUN (for não)

b. Both forms of the following are written out:

acessível	ACESSI'VEL
accessível	ACCESSI'VEL
aeroporto	AEROPORTO
aereoporto	AEREOPORTO
contato	CONTATO
contacto	CONTACTO
ínterim	I'NTERIM
interim	INTERIM
miligrama	MILIGRAMA
miligramo	MILIGRAMO
questão	QUESTA+O
qüestão	QUESTA+O9
queto	QUETO
quieto	QUIETO

c. Certain nicknames which are homographs of other words are followed by the ▽ symbol.

DADO▽	TE'▽
VI▽	VIVI▽

d. The names of the letters A, E, and O are followed by the $ symbol to distinguish them from articles and prepositions.

13. Divergent forms and usages.

a. Internal errors in pronounciation will be typed as pronounced with the correct form preceding and separated from the incorrect form with a closed parenthesis, all in one interspace. If the error involves omitted letters at the end of a word, the letters are added preceded by an open parenthesis.

probrema	PROBLEMA)PROBREMA
sa (sabe)	SA(BE

b. Because of the difficulty of recognition of high frequency sounds, the omission of the final s will not be noted.

c. For errors in syntax, the symbol $ will be typed before the error, followed by the space bar.

a problema	/$/A/PROBLEMA/

d. For the sake of consistency the contraction of the preposition a and the feminine definite article will be assumed wherever this interpretation is possible.

à sua disposição	/A#/SUA/DISPOSIC,A+O/
à Maria	/A#/MARIA/

14. Non-identity.

a. Persons and places not to be identified are shown as XXX . Two such names occurring together are written:

XXX▽XXX

b. An unidentifiable word or phrase in context is shown by:

/ZZZ/

15. Considerations for optical scanning.

a. All spaces are identified by the / slash mark.

vou ser o chefe	/VOU/SER/O/CHEFE/
vou ajudá-la	/VOU/AJUDA'-/LA/

b. All incomplete words will be broken at the end of the line, without a hyphen or space.

c. A blob symbol ▮ must precede all low level characters appearing in the first position of a line.

Low level characters:

11. # 12. = 27. ' 28. , 32. + 48. - 59. .

d. Dialogs may be carried over from one page to the next. The first line on every page gives the pertinent information for the hard copy only (tape number, date, source, etc.) and carries the delete symbols. The informant number on the following page does not appear until the informant changes.

e. The line delete symbol at the end of the line is]]] . At the beginning of a line deletion is accomplished by striking over the first three characters with the upper case N letter.

~~RIO~~/DE/JANEIRO

f. The first line on each page must be at least seventeen characters long.

g. The Δ delta symbol is used only for colloquy headings and precedes the informant number. It is located on the line immediately above a new paragraph.

ΔΔ24FS862

16. Computer symbols.

a. The following are used for sorting and do not print:

31. $ 47. & 58. Δ 62. ▽

b. The following symbols are not used:

26. ¬ 29. ⌐ 42. ⊢

c. The period, comma, and question mark will print, but will not sort nor will they be listed in the KWIC.

SPECIAL PROVISIONS FOR TYPING LITERARY PORTUGUESE

1. Every effort will be made to reproduce as closely as possible the language used by the various authors. For this reason it is necessary to add certain symbols:

Exclamation mark	/'/
Quotation	/"/
Change of speaker	/-/

These characters will be separated by spaces from other words. Also, to conform to the original texts, the colon and semi-colon will be used. : ;

2. In the event the original text has a misspelled word or typographical error, the correct form only will be typed.

3. The first line on each page will contain the author's name, title of the work, publishing house, and the date of the edition being used.

4. Informant numbers will have first the delta Δ symbol, then the three-character page number of the original text, next the two letter code for the author, and finally the three-character page number of the hard copy. Only the two letter author code will be used to obtain the speaker list - the number of different "speakers", in this case authors, who use each individual word of the corpus. Δ011AF001

5. The paragraph symbol will be used to indicate a new paragraph. It will not sort for a KWIC listing, but it will print.

6. Future and Conditional forms:

dir-se-ão	DIR-/SEƱ/-A+0
encontrar-se-iam	ENCONTRAR-/SEƱ/-IAM/

7. Special forms:

D. Maria	/D./MARIA/
1º	/1.0/ (not zero)
1920	/1ƱƱ0/
V.Ex.a	/V.ƱEX.A/
Nº 1	/N.0/1/
a fala (speech)	/A/FALAƱ/
nós (knots)	/NO'SƱ/
Dr. Ruas	/DR./RUASƱ/

```
TAPE 33A SAO PAULO SP 18 MARCH 1968 008FH-U 009MC 010MC 011FH YTJJJ
PAGE 139JJJ
LEMA/NENHUM■/,/AH▽/▽.../EU/FALEI/COM/MEU/ME'DICO/ANTES/,/NA+O/TEVE/PROBL
EMA/,/MAS/RONALDO■/TEVE/RUBE'OLA/./AH▽/▽.../E=LE/JA'/ESTA'/BOM/,/PASSOU/
TUDO/IH▽/▽.../APESAR/DE▽DE■/TERMOS/TIDO/UM■/GRANDE/SUSTO/./MAS/NA/MESMA/
HORA■/EU/LIGUEI/PARA/O/MEU/ME'DICO■/E/E=LE/DISSE/QUE/NA+O/TINHA/PROBLEMA
/NENHUM/,/MESMO/QUE/EU/NA+O/TIVESSE/TIDO/,/QUE/NA+O/TERIA/PROBLEMA/,/POR
QUE/NO/SE'TIMO/ME=S/NA+O/TEM/MAIS/PROBLEMA/NENHUM/,/NENHUM/./MAS■/MESMO/
ASSIM/NO'S/FICAMOS/MEIO/CHATEADOS/E/PREOCUPADOS/,/MAS/DEPOIS/FALEI/COM/O
UTROS/ME'DICOS/E/TODOS■/DISSERAM/A/MESMA/COISA/./AINDA■/MAIS/QUE/EU/JA'/
TIVE/E/NA+O/TEM/ASSIM■/PROBLEMA/MESMO/./E=LE/FICOU/A/SEMANA/INTEIRA/EM/C
ASA/IH▽/▽.../(ESTO-/ESTOUROU/DOMIN-/NO/OUTRO/DOMINGO/,/QUER/DIZER/,/HOJE
/JA'■/FAZ/MAIS/DE/UMA/SEMANA/./E/ACHO/QUE/O/ME'DICO/RECOMENDOU/QUE■/E=LE
/FO=SSE/TRABALHAR/TERC,A-/OU/QUARTA-/SO'/,/AMANHA+/OU/DEPOIS■/./AH▽/▽...
/FICOU/CHEIO/DE/FICAR/EM■/CASA/,/DETESTA■/FICAR/,/E/PRINCIPALMENTE/SEM/P
ODER/LER/NEM/FAZER/NADA/PORQUE/A/VISTA/ES)TAVA/MUITO/IRRITADA/./MAS/FORA
/ISSO/NA+O/DEU/NENHUMA/OUTRA/COMPLICAC,A+O/,/NA+O/TEVE/PROBLEMA■/NENHUM/
■./NO'S/JA'/ESTA'VAMOS/MAIS/OU/MENOS/ESPERANDO/,/PORQUE/ESTA'/UMA/EPIDEM
IA/INCRI'VEL/DE/RUBE'OLA/AQUI/,/TA'/TODO/MUNDO/COM/RUBE'OLA/./E/NO'S/JA'
/ES)TA'VAMOS/IMAGINANDO/QUE/▽.../QUE/E=LE/NA+O/IRIA/SE▽/LIVRAR/DESSA/./M
AS■/NUN/TEM/MAIS/PROBLEMA/NENHUM/,/NUN/SE▽/PREOCUPEM/,/TA'/TUDO/EM/ORDEM
/./EH▽/▽.../DAISY/,/COMO■/E'/QUE/VOCE=/ESTA'/,/JA'/ENGORDOU/ALGUMA/COISA
/?/CE=/TA'■/PASSANDO/BEM/?/ADOREI/SABER/QUE/CE=/,/QUE/TA'/TUDO/EM/ORDEM/
E■/TAL/./AH▽/▽.../EU/(RE-/RECEBI/A/SUA/CARTA/E/ESSA■/QUE/EU/MANDEI/A/SEM
ANA/PASSADA/FOI■/RESPONDENDO/A/SUA/,/IH▽/▽.../FALANDO/SO=BRE/AS/COISAS/D
O/BEBE=/E/TAL■/QUE/ALIA'S/E'/O/ASSUNTO/DA/MODA/PRA/NO'S/,/NUN/E'/?/CARLO
S▽EDUARDO/,/COMO/E'/QUE/VA+O/AS/COISAS/POR/AI'■/?/O/TRABALHO/E/OS/PREPAR
ATIVOS/DE/VOCE=S/TAMBE'M/?/ACHEI/ESPETACULAR■/E/(VO-/ESPETACULAR/VOCE=S/
IREM/A▽A&/MA'LAGA/E/A#/ALDEIA/DO/VOVO=/E/VISITAR/A/FAMI'LIA/TO=DA/./ACHE
I■/UMA/IDE'IA/O'TIMA/./MORRI/DE/INVEJA■/,/ACHEI/MUITO/BACANA/./AH▽/▽.../
AH▽/▽.../BOM/,/VAMOS/VER/SE/CE=S/OUVIRAM/./CARLOS▽EDUARDO/E/DAISY/FALEM/
UM/POUCO/,/PRA/DEPOIS/EU/FALAR■/MAIS/./
Δ010MC139
MUITO/BEM/ESCUTADO/./CEM/POR/CENTO/,/BEATRIZ/,/O#TIMAMENTE/ESCUTADO/E/QU
E/SUSTO/E=SSE/NEGO'CIO/DE/DOENC,A/,/QUE/NA+O/SABI'AMOS/NADA/./NA+O/CHEGO
U■/A/SUA/CARTA/DA/U'LTIMA/SEMANA/,/NA+O/./TALVEZ/▽.../TALVEZ/,/ENTA+O/,/
CHEGUE/AMANHA+/./ENTA+O/EU/QUERO/QUE▽QUE/O/MICROFONE/VA'/PARA/O/CARLOS▽E
DUARDO/./■
Δ009MC139
E'/SO'/EU/AMEAC,AR/DE■/FALAR/COMEC,A/A&/DAR/CRAMPE/AQUI/./FALAM/TANTO/QU
E■/QUANDO/EU/COMEC,O/A&/FALAR/DA'/CRAMPE/./BEM/JA'/QUE/A/MINHA/VOZ■/E'/Q
UE/ESTA'/SENDO/(PRE-/PRECISA/SER/OUVIDA/AI'/,/EU/QUE/NA+O/FALO/,/EU/QUE/
SOU/CHATO/,/VOU/FALAR/POR/TODO/MUNDO/HOJE/AQUI/./A/DAISY/TA'/BOA■/,/PASS
ANDO/MUITO/BEM/,/TA'■/O'TIMA/,/NA+O/ENGORDOU/,/TA'/DISFARC,ANDO/DIREITIN
HO■/AINDA/IH▽/▽.../VAI/INDO/MUITO/BEM/./RECEBEU/O/CARTA+O/DE/ANIVERSA'RI
O/,/TA'/AGRADECENDO■/,/MANDANDO/UM/ABRAC,O/PRA/VOCE=/./SABE/QUE/EU/ESTOU
/(CAN-/FICO/QUIETINHO/AQUI/MAS/ESTOU/OH▽/▽.../LENDO/TO=DAS/AS/SUAS/CARTA
S/COM■/MUITO/PRAZER/,/MUITO/GOSTOSO/,/IH▽/▽.../APROVEITANDO/,/SABENDO/TO
■=DAS/AS/NOTI'CIAS/E/ACOMPANHANDO/TUDO/./TAMBE'M/ESTOU/COM/MUITAS/SAUDAD
```

TYPICAL PAGE OF SPOKEN CORPUS FOR OPTICAL SCANNING

40

```
RUBEM BRAGA. AI DE TI, COPACABANA. RIO. EDITORA DO AUTOR. PP 37-67]]]
1960, PAGE 049]]]
Δ037RB049
UM/TELEFONEMA/■APENAS/CORDIAL/,/A&/QUE/ATENDO■/COM/NATURALIDADE/-/MAS/PO
RQUE/,/DEPOIS/,/E=SSE/INDEFINI'VEL/TREMOR/I'NTIMO■/,/ESSA/REMOTA/NOC,A+O
/DE/QUE■/REPRESENTEI/UMA/CENA/SOB■/O/EFEITO/DO/HIPNOTISMO/,/E=SSE/INDIZI
■'VEL/SUSTO■/?/SOU/UM/HOMEM/TRANQUILO/,/E/MINHA/VIDA■/ESTA'/TRANQUILA/;/
OUC,O/ESSA/VOZ/,/E=SSE/NOME/,/E■/PRONTO/''/-/COMEC,O■/A&/AGIR/COMO/SE/EU■
/TRABALHASSE■/EM/UM/FILME■/A&/QUE/EU/MESMO■/ESTIVESSE■/ASSISTINDO/./REPR
ESENTO/MEU/PAPEL■/DE/MANEIRA■/NORMAL/E/FAC,O■/O/PAPEL/DE/UM/HOMEM/NORMAL
/;/MAS/HA'■/UM/OUTRO/EU/INVISI'VEL■/QUE/E'/AQUALOUCO/,/PATINADOR■/SO=BRE
/ARCO-I'RIS■/,/MENINO/TONTO/,/HAMLET/,/PALERMA/,/PATE'TICO/./ENQUANTO/EU
/DIGO/UMA/COISA/SENSATA/E=SSE/MEU/FANTASMA/SE∇/ENTREGA/A&/UM/SILENCIOSO/
DESVARIO/,/OU/RECITA■/VERSOS/ANTIGOS/,/VOA/COMO/UM/ANJO/,/SOLUC,A/./POSS
O/CONTEMPLA'-/LO/COM/FRIEZA/,/CRITICA'-/LO/,/TER/PENA/DE=LE■/;/EVITO/QUE
/E=LE■/INFLUA/NO/MAIS/MI'NIMO■/EM/MINHA/CONDUTA■/REAL■/;/QUANDO/E=LE/TEM
/UM/IMPULSO/DE/FALAR/AO/TELEFONE/EU/ME∇/PONHO/TRANQUILAMENTE/A&/DESCASCA
R■/UMA/LARANJA■/OU/FAZER/PONTA/EM/UM/LA'PIS■/;/E/SEM/MINHAS■/MA+OS/,/SEM
/MEU■/CORPO/,/E=LE■/NA+O/PODE/FAZER/NADA/./RESOLVO/IGNORA'-/LO/E/CHEGO■/
A&/ESQUECE=-/LO/DURANTE/SEMANAS/,/MESES/;/MAS/QUANDO/SURGE/A/PRESENC,A∇/
E=LE■/SALTA/AO/MEU■/LADO/,/SOB/UMA/LUZ/SOBRENATURAL■/,/ABSURDO/E/INFANTI
L/./¶/
Δ038RB049
NA+O/ESTOU■/APAIXONADO■/;/MEU/COME'RCIO■/SENTIMENTAL/COM/AS/OUTRAS/CRIAT
URAS/CORRE/NORMAL■/,/COM/SUAS/ALEGRIAS■/E/TRISTEZAS■/./NA+O/ESTOU/APAIXO
NADO■/,/MAS/POSSO/VER/A/FACE/DA/PAIXA+O/./E/POR/UM/INSTANTE/FICO■/PARADO
/,/MUDO/,/COMO/QUEM/OUVISSE/,/NO/FUNDO/DA/NOITE/,/O/SUSSURRO/DAS/ESTRE=L
AS/,/E/O∇/RECONHECESSE/./¶/
Δ039RB050
ANTO=NIO∇MARIA■/■CONTOU/QUE/UMA/VEZ/IA/NUM/TA'XI/GUIADO■/POR/UM/CHOFER■/
PORTUGUE=S■/VELHO/,/BIGODUDO■/,/CALADO/,/DE/CARA/TRISTE/./QUANDO/O/CARRO
/CHEGOU■/A#/PRAIA/O/CHOFER/VIU∇/UM/BARCO/E/EXCLAMOU■/,/APONTANDO/COM/O/B
RAC,O■/ESTICADO■/,/OS/OLHOS/BRILHANTES/,/NUM/TOM/DE/DESCOBERTA■/,/DESAFI
O■/E/ALEGRIA/:/¶/-/OLHA/O/NAVIO/PEQUENINO■/''/¶/ESSA/FASCINAC,A+O/DOS/POR
TUGUE=SES/PELOS/NAVIOS/ME/SALVOU/A/TARDE/DE/ONTEM/./EU/TINHA■/DE/IR/A#/A
LFA=NDEGA/E/,/PORTANTO/,/PASSAR/PELA/PRAC,A■/MAUA'/./O/PORTUGUE=S■/DO/VO
LANTE■/VINHA/PRAGUEJANDO/CONTRA■/O/CALOR/,/CONTRA/OS/OUTROS/CARROS/,/CON
TRA■/TUDO/./ANTES/DE=LE/EU/VI/O/"/VERA∇CRUZ/"/ENCOSTADO/NO/CAIS/,/E/DISS
E■/:/"/OLHE■/O/VERA∇CRUZ/"/,/QUE/NAVIO/BONITO/''/"/E=LE/■RECEBEU/ISSO/COM
O■/UM/ELOGIO/PESSOAL/E/COMEC,OU/A&/FALAR/DO/NAVIO/COM/ENTUSIASMO/,/ATE'/
CONHECIA/UM/MAQUINISTA/DE/BORDO■/E/VISITARA/TODO/O/GIGANTE■/:/"/TEM/OITO
/ANDARES/,/MAS/TEM■/ELEVADOR/''/"/¶/
Δ040RB049
PELAS■/CINCO/E/POUCO/,/AO/VOLTAR/PARA/CASA/,/ME/TOCOU/OUTRO/VOLANTE/PORT
UGUE=S■/./NA/ALTURA/DO/FLAMENGO/DIVISEI/O/NAVIO■/,/QUE/MARCHAVA/PARA/A/S
AI'DA/DA/BARRA/,/E/RESOLVI/ELOGIAR/NOVAMENTE/O/BARCO■/,/PARA/VER/O/EFEIT
O■/./FOI∇/MARAVILHOSO/./■"/E'/REALMENTE/,/E'/REALMENTE/,/E'/UM/BELO/NAVI
O■/''/"/FIZ/NOTAR/QUE/O/BRASIL/NA+O/TINHA/NENHUM/NAVIO/DE/PASSAGEIROS/TA+
O/GRANDE/E/TA+O/BONITO■/,/E/ISSO/ANIMOU■/AINDA/MAIS/O/HOMEM/./ACABOU/CON
```

TYPICAL PAGE OF LITERARY CORPUS FOR OPTICAL SCANNING

[illegible]	LILIANA , EU MANDEI UMA CARTA ENORME , CONTANDO TUDO COMO FOI	MANDEI
482FF264	ER A PRESENC,A AI' . EU ESTOU BEM , TUDO O'TIMO , MANDEI UMA CARTA ENORME PARA VOCE= E ONTEM BOTEI	MANDEI
165FH283	TENHO TENTADO FALAR COM VOCE= PELO TELEFONE . EU MANDEI UMA CARTA EXPRESSA QUE DEVE CHEGAR NA TERC	MANDEI
551FH272	M AS CRIANC,AS , OS RETRATOS ESTAVAM LINDOS . EU MANDEI UMA CARTA HA' DIAS NUN SEI SE VOCE= RECEBE	MANDEI
863ME325	UMA CARTA PARA VOCE= , A MA+E ESCREVEU HOJE E EU MANDEI UMA CARTA HOJE PARA A MIMI TAMBE'M , ENVIE	MANDEI
258FW509	NA+O . NO'S NUN VAMOS PRA NOVA IORQUE , NA+O . EU MANDEI UMA CARTA HOJE PRA VOCE= . RECEBI HOJE UMA	MANDEI
832AN111	SA IH ... TUDO BEM . ME AVISA AO HAROLDO QUE EU MANDEI UMA CARTA IMPORTANTE PARA E=LE ANTEONTEM .	MANDEI
173FH889	MAMA+E , TA' O'TIMO , ESCUTEI TUDO . EU MANDEI UMA CARTA ONTEM . HELENA TA' COM QUATORZE	MANDEI
763MN906	O CONTRATO AH E=LE FARA' A MUDANC,A . E=LE PEDIU MANDEI UMA CARTA ONTEM OU ANTEONTEM , E=LE PEDIU	MANDEI
346FH524	E MANHA+ E ... DEVE IR LOGO PRA NOVA IORQUE . EU MANDEI UMA CARTA PARA A CAIXA POSTAL CUJO NU'MERO	MANDEI
261FH339	MEIO ATRAPALHADA QUANDO VOCE=S LIGAM , ENTA+O EU MANDEI UMA CARTA PARA A SOLANGE PORQUE EU ESTIVE	MANDEI
684FH282	+O E' COISA GRAVE , NA+O E' NADA GRAVE . E EU JA' MANDEI UMA CARTA PARA VOCE= , NERI TAMBE'M , NERI	MANDEI
155FH276	OU PRA NO'S , NO'S AGRADECEMOS MUITO . AGORA , EU MANDEI UMA CARTA PELA TERESA QUE EU NA+O OBTIVE R	MANDEI
669FH860	UMA CARTA BEM DETALHADA EXPLICANDO TUDO E HOJE EU MANDEI UMA CARTA POR UM TRIPULANTE DA VARIG . A	MANDEI
286FF385	SUZY , A SUZY ESTUDARIA AQUI , ENTENDEU ? HOJE EU MANDEI UMA CARTA PRA AI' . OUTRA COISA QUE EU QUE	MANDEI
304FS482	QUE HA' ? OLHA , EU ... DIGA A# ESTELINHA QUE EU MANDEI UMA CARTA PRA ELA EXPLICANDO O NEGO'CIO DA	MANDEI
372MS239	, NA+O TEM NADA NA+O . QUE E=LE ESCREVA . QUE EU MANDEI UMA CARTA PRA IBE E PRO AH ... E PRA ..	MANDEI
346FH524	TIMO FOI DE DE DE LOS ANGELES E MANDAMOS ... EU MANDEI UMA CARTA PRA O YMCA E JA' DEVE TER CHEGAD	MANDEI
212FH326	EU VOU TER QUE DEVOLVER . EH ... (SE EU HOJE EU MANDEI UMA CARTA PRA SENHORA . AQUI TA' TUDO BEM	MANDEI
7[illegible]MF896	MUITOS BEIJOS E TE=RC,A-FEIRA ... SEGUNDA-FEIRA MANDEI UMA CARTA PRA SENHORA , SEGUIRA'H EH ...	MANDEI
279FH450	M NADA . ENTA+O NO'S ESTAMOS ESPERANDO E HOJE EU MANDEI UMA CARTA PRA SENHORA , VIU ? UM BEIJO MUI	MANDEI
286FF385	PASSAPORTE PRONTO , A SENHORA E A SUZY . EU HOJE MANDEI UMA CARTA PRA SENHORA SUGERINDO QUE A SENH	MANDEI
587FD911	TRABALHANDO BEM , CONTENTE , (ONT- HOJE DE MANHA+ MANDEI UMA CARTA PRA VOCE= , AINDA ESCREVI DE NO=	MANDEI
121FA462	VAMOS CORTAR QUE A PROSA TA' MUITO GRANDE . HOJE MANDEI UMA CARTA PRA VOCE= , CE= RECEBE DAQUI A	MANDEI
166MN284	RMA DA CARTA CHEGAR SEGURO AQUI NO RIO . ONTEM EU MANDEI UMA CARTA PRA VOCE= , COMO JA' MANDEI NA S	MANDEI
584FH833	O= DINA . EU TOU DEPENDENDO EU MANDEI UMA CARTA PRA VOCE= , REGISTRADA . EU ESTO	MANDEI
584FH833	TIGO PRA TE EXPLICAR TUDO ISSO , COMPREENDEU ? EU MANDEI UMA CARTA PRA VOCE= , VOCE= DEVERA' RECEBE	MANDEI

TYPICAL PAGE OF KWIC PRINTOUT - REDUCED SIZE

SORT

```
9000 REM PROGRAM-- SORT***
9001 REM
9002 REM WRITTEN BY-- DP3 WILLIAM TAKACS  OCTOBER 1,1971
9003 REM
9004 REM DESCRIPTION-- THIS PROGRAM WILL SORT ANY FILE OF ASCII
9005 REM                CHARACTERS INTO EITHER ASCENDING OR DESCENDING
9006 REM                ORDER ACCORDING TO ANY SIZE OR NUMBER OF
9007 REM                CHARACTER FIELDS.  THERE IS MAXIMUM OF
9008 REM                FOUR INPUT FILES IN EACH SORT RUN.
9009 REM
9010 REM INSTRUCTIONS--SAVE AN EMPTY FILE FOR THE SORTED DATA.
9011 REM
9012 REM                ANSWER THE QUESTIONS AS
9013 REM                THEY ARE ASKED BY THE COMPUTER.
9014 REM - - - - - - - - - MAIN  PROGRAM - - - - - - - - - - - - - -
9015 LET Q2$="SORTEND***"
9016 FILE #6:"**"
9017 DIM Q(100)
9018 PRINT "HOW MANY INPUT FILES";
9019 INPUT Q0
9020 PRINT
9021 IF Q0.4 THEN 9086
9022 FOR Q5=1 TO Q0
9023 PRINT "WHAT IS INPUT FILENAME # ";Q5;
9024 INPUT Q1$
9025 FILE #Q5:Q1$
9026 NEXT Q5
9027 PRINT "WHAT FILE DO YOU WANT THE SORTED DATA WRITTEN INTO";
9028 INPUT Q1$
9029 PRINT
9030 FILE #5:Q1$
9031 LET Q9=1
9032 PRINT "DO YOU WANT TO SORT INTO 'ASCENDING' OR 'DESCENDING' ORDER";
9033 INPUT Q1$
9034 PRINT
9035 CHANGE Q1$ TO Q
9036 IF Q(1)=65 THEN 9038
9037 LET Q9=-1
9038 PRINT "HOW MANY FIELDS DO YOU WANT TO SORT ON";
9039 INPUT Q4
9040 PRINT
9041 WRITE #6: Q0,0,Q4,1,1,1,0
9042 PRINT "FOR EACH SORT FIELD-- STARTING WITH THE FIRST FIELD"
9043 PRINT "TO BE SORTED ON, THEN THE NEXT, AND SO ON -- TYPE THE"
9044 PRINT "CHARACTER POSITIONS OF THE SORT FIELD (E.G., TO SPECIFY A"
9045 PRINT "SORT FIELD OF CHARACTERS 3,4,5,&6, YOU TYPE--. 3-6)"
9046 PRINT
9047 FOR Q5=1 TO Q4
9048 PRINT "SORT FIELD # ";Q5;
9049 LINPUT Q1$
```

SORT (continued)

```
9050 CHANGE Q1$ TO Q
9051 LET Q7=-1
9052 LET Q8=0
9053 FOR Q2=1 TO Q(0)
9054 IF Q(Q2)=32 THEN 9060
9055 IF Q(Q2)=45 THEN 9063
9056 LET Q6=Q(Q2)-48
9057 IF Q6,0 THEN 9076
9058 IF Q6.9 THEN 9076
9059 LET Q8=10*Q8+Q6
9060 NEXT Q2
9061 REM
9062 GOTO 9066
9063 LET Q7=Q8
9064 LET Q8=0
9065 GOTO 9060
9066 IF Q7.-1 THEN 9068
9067 LET Q7=Q8
9068 IF Q7.Q8 THEN 9083
9069 WRITE #6: 1,Q7,Q8,Q9,0,0,0,0
9070 NEXT Q5
9071 ON  Q0 GOTO 9072,9073,9074,9075
9072 CHAIN Q2$ SYSTEM "SORT" WITH #6,#1,#5
9073 CHAIN Q2$ SYSTEM "SORT" WITH #6,#1,#2,#5
9074 CHAIN Q2$ SYSTEM "SORT" WITH #6,#1,#2,#3,#5
9075 CHAIN Q2$ SYSTEM "SORT" WITH #6,#1,#2,#3,#4,#5
9076 IF Q(Q2)=44 THEN 9080
9077 PRINT " A RANGE OF WHOLE NUMBERS IS THE ONLY VALID RESPONSE "
9078 PRINT " TO THIS QUESTION.  RE-TYPE YOUR RESPONSE CORRECTLY."
9079 GOTO 9048
9080 PRINT " PLEASE TYPE ONLY ONE RANGE OF NUMBERS PER FIELD."
9081 PRINT " RE-TYPE YOUR RESPONSE CORRECTLY."
9082 GOTO 9048
9083 PRINT " THE RANGE OF CHARACTERS MUST BE SPECIFIED IN ASCENDING"
9084 PRINT " ORDER.  PLEASE RE-TYPE YOUR RESPONSE CORRECTLY."
9085 GOTO 9048
9086 PRINT "THIS PROGRAM ALLOWS FOR A MAXIMUM OF FOUR INPUT FILES."
9087 PRINT "CONTACT THE PROGRAMMING ASSISTANT AT EXT. 2185 FOR"
9088 PRINT "INSTRUCTIONS ON SORTING DATA CONTAINED IN MORE THAN"
9089 PRINT "FOUR FILES."
9090 END
```

SORTEND

```
10 REM PROGRAM -- SORTEND***
15 REM
20 REM DESCRIPTION-- THIS PROGRAM DOES THE WRAP-UP PROCEDURES
25 REM                FOR THE BASIC PROGRAM SORT***.
40 REM
60 PRINT
65 END
```

LOCATE

```
100 '
110 ' ----------------Description-----------------------------
120 '
130 '     Written by Carl Tannenbaum and Edward Rippon for
140 '     Prof. J.A. Hutchins.  This program will locate unique
150 '     strings in any file, transferring them to a separate
160 '     file named LOCA which must be SAVED before running
170 '     the program.
180 '
190 ' ------------------PROGRAM-----------------------------------
200 DIM Z(100)
210 PRINT "INPUT FILE";
220 INPUT F$
230 FILE #1:F$
240 FILE #2:"LOCA"
250 SCRATCH #2
260 MARGIN #2:130
270 PRINT "ARE YOU USING A REGULAR TERMINAL";
280 INPUT C$
290 IF C$ = "YES" THEN 320
300 MARGIN 130
310 GO TO 330
320 MARGIN 70
330 LET Z1=1
340 PRINT "WHICH STRING DO YOU WISH TO LOCATE"
350 PRINT
360 INPUT A$
370 LET C=0
380 RESET #1
390 IF END #1 THEN 510
400 LINPUT #1:B$
410 LET N=1
420 LET S = POS(B$,A$,N)
430 IF S=0 THEN 390
440 IF S-1+LEN(A$).LEN(B$) THEN 390
450 IF A$ ,. SEG$(B$,S,S+LEN(A$)-1) THEN 490
460 PRINT #2 : B$
470 LET C = C+1
480 GO TO 390
490 LET N = S + 1
500 GO TO 420
510 PRINT
520 PRINT
530 PRINT "THERE WERE";C;"MATCHES FOUND."
540 PRINT
550 PRINT
560 PRINT #2:
570 PRINT #2:
580 RESET #2
590 LET Z(Z1)=C + 2
```

LOCATE (continued)

```
600 LET T9=0
610 IF Z1=1 THEN 680
620 FOR I = 1 TO (Z1-1)
630 LET T9=T9+Z(I)
640 NEXT I
650 FOR J = 1 TO T9
660 LINPUT #2:D$
670 NEXT J
680 FOR K = 1 TO Z(Z1)
690 LINPUT #2:A$
700 PRINT A$
710 NEXT K
720 LET Z1=Z1+1
730 PRINT
740 PRINT "DO YOU WISH TO LOCATE ANOTHER STRING";
750 INPUT Y$
760 IF Y$="YES" THEN 340
770 END
```

MULTIMUV

```
100 REM   WRITTEN BY J.W. SCHWAB FOR PROF. JOHN A. HUTCHINS
110 REM
120 REM
130 REM   DESCRIPTION- THIS PROGRAM WILL MOVE ANY NUMBER OF
140 REM                VERTICAL COLUMNS TO ANY DESIRED POSITION.
150 REM                THE COLUMNS OF THE INPUT FILE MUST BE
160 REM                SEPARATED BY THE CHARACTER "±".  THE
170 REM                MARGIN IS SET FOR THE IBM 2741, AND
180 REM                SHOULD BE CHANGED IF THIS PROGRAM IS
190 REM                RUN ON A STANDARD TELETYPE TERMINAL.
200 REM
210 REM
220 REM   ----------------------PROGRAM-------------------------
230 DIM I$(130),J$(130),F(130),O(130),P(130),T(130)
240 PRINT "WHICH FILE DO YOU WANT REARRANGED";
250 LINPUT A$
260 PRINT "IN WHAT FILE DO YOU WANT RESULTS TO BE STORED";
270 LINPUT B$
280 FILE #1:A$
290 FILE #2:B$
300 SCRATCH #2
310 MARGIN #2:130
320 PRINT "HOW MANY VERTICAL COLUMNS (FIELDS), SEPARATED BY"
330 PRINT "THE (±), ARE THERE IN THE DATA FILE";
340 INPUT N
350 PRINT "IDENTIFY THE";N;"FIELDS FROM LEFT TO RIGHT."
360 FOR B=1 TO N
370 INPUT I$(B),
380 NEXT B
390 PRINT "IDENTIFY WHICH COLUMN YOU WANT ON THE LEFT,NEXT,NEXT,"
400 PRINT "NEXT,.....ETC.  USE EXACT TITLES FOR EACH VARIABLE AND"
410 PRINT "AFTER EACH,FOLLOWED BY A COMMA,INDICATE THE TAB MARKER"
420 PRINT "FOR EACH COLUMN."
430 FOR B=1 TO N
440 INPUT J$(B),T(B),
450 NEXT B
460 LET C=1
470 FOR B=1 TO N
480 IF I$(B),.J$(C) THEN 530
490 LET F(C)=B
500 LET O(C)=C
510 LET C=C+1
520 GO TO 470
530 NEXT B
540 FOR B=1 TO N
550 NEXT B
560 LINPUT #1:A$
570 LET E=1
580 FOR B=1 TO N-1
590 LET P(B)=POS (A$,"±",E)
```

MULTIMUV (continued)

```
600 LET E=P(B)+2
610 NEXT B
620 LET P(B+1)=LEN(A$)+2
630 LET C=1
640 FOR B=1 TO N
650 LET B$(B)=SEG$(A$,C,P(B)-1)
660 LET C=2+P(B)
670 NEXT B
680 FOR B=1 TO N
690 PRINT #2:TAB(T(B));B$(F(B));
700 NEXT B
710 PRINT #2:
720 IF MORE #1 THEN 560
730 PRINT
740 PRINT
750 PRINT "SUPER-MOVER HAS COMPLETED ITS TASK↓↓↓↓↓↓↓↓"
760 END
```

```
NOTAB

100 REM    WRITTEN BY J.W. SCHWAB FOR PROF. JOHN A. HUTCHINS
110 REM
120 REM
130 REM    DESCRIPTION- THIS PROGRAM WILL MOVE ANY NUMBER OF
140 REM                 VERTICAL COLUMNS TO ANY DESIRED POSITION.
150 REM                 THE COLUMNS OF THE INPUT FILE MUST BE
160 REM                 SEPARATED BY THE CHARACTER "±".  THE
170 REM                 MARGIN IS SET FOR THE IBM 2741, AND
180 REM                 SHOULD BE CHANGED IF THIS PROGRAM IS
190 REM                 RUN ON A STANDARD TELETYPE TERMINAL.
200 REM                 TO INDICATE TAB POSITIONS FOR EACH
210 REM                 COLUMN, USE MULTIMUV PROGRAM.
220 REM
230 REM
240 REM    ----------------------PROGRAM-----------------------
250 DIM I$(130),J$(130),F(130),O(130),P(130)
260 PRINT "WHICH FILE DO YOU WANT REARRANGED";
270 LINPUT A$
280 PRINT "IN WHAT FILE DO YOU WANT RESULTS TO BE STORED";
290 LINPUT B$
300 FILE #1:A$
310 FILE #2:B$
320 SCRATCH #2
330 MARGIN #2:130
340 PRINT "HOW MANY VERTICAL COLUMNS (FIELDS), SEPARATED BY"
350 PRINT "THE (±), ARE THERE IN THE DATA FILE";
360 INPUT N
370 PRINT "IDENTIFY THE";N;"FIELDS FROM LEFT TO RIGHT."
380 FOR B=1 TO N
390 INPUT I$(B),
400 NEXT B
410 PRINT "IDENTIFY WHICH COLUMN YOU WANT ON THE LEFT,NEXT,NEXT,"
420 PRINT "NEXT,.....ETC.  USE EXACT TITLES FOR EACH VARIABLE."
430 FOR B=1 TO N
440 INPUT J$(B),
450 NEXT B
460 LET C=1
470 FOR B=1 TO N
480 IF I$(B),.J$(C) THEN 530
490 LET F(C)=B
500 LET O(C)=C
510 LET C=C+1
520 GO TO 470
530 NEXT B
540 FOR B=1 TO N
550 NEXT B
560 LINPUT #1:A$
570 LET E=1
580 FOR B=1 TO N-1
590 LET P(B)=POS (A$,"±",E)
```

NOTAB (continued)

```
600 LET E=P(B)+2
610 NEXT B
620 LET P(B+1)=LEN(A$)+2
630 LET C=1
640 FOR B=1 TO N
650 LET B$(B)=SEG$(A$,C,P(B)-1)
660 LET C=2+P(B)
670 NEXT B
680 FOR B=1 TO N
690 PRINT #2:B$(F(B));
700 NEXT B
710 PRINT #2:
720 IF MORE #1 THEN 560
730 PRINT
740 PRINT
750 PRINT "SUPER-MOVER HAS COMPLETED IT'S TASK½½½½½½½"
760 END
```

RITEJUST

```
100 REM:   Written by Carl Tannenbaum
110 REM:
120 REM:  DESCRIPTION:  This program will right justify
130 REM:                a column of numbers that is left
140 REM:                justified, provided that spaces
150 REM:                precede and follow the column.
160 REM:                Tabs positions can be the same as
170 REM:                those used in MULTIMUV.  The relative
180 REM:                position of the column does not
190 REM:                change between input and output file.
200 REM:
210 REM: ---------------MAIN PROGRAM -------------------
220 REM:
230 PRINT "WHAT FILE HAS THE COLUMN YOU WANT RIGHT JUSTIFIED"
240 INPUT A$
250 FILE #1:A$
260 PRINT "IN WHICH FILE ARE THE RESULTS TO BE STORED";
270 INPUT B$
280 FILE #2:B$
290 SCRATCH #2
300 MARGIN #2:130
301 PRINT
310 PRINT "TAB POSITION WHERE LEFT JUSTIFIED NUMBER COLUMN BEGINS"
320 INPUT C
330 PRINT "Tab position where next column to the right begins";
340 PRINT
350 INPUT C2
360 LET C=C+1
370 PRINT "Tab position immediately after largest left justified"
380 PRINT "number";
390 INPUT D
400 LET D = D + 1
410 LINPUT #1:A$

430 LET P=POS(A$," ",C)
440 LET N$=SEG$(A$,C,P-1)
450 FOR I=1 TO D-C-LEN(N$)
460 LET N$=" "&N$
470 NEXT I
480 PRINT #2: SEG$(A$,1,C-1);N$;TAB(C2-1);SEG$(A$,C2,LEN(A$))
490 IF MORE #1 THEN 410
500 PRINT
510 PRINT "Now go to output file"
520 END
```

DESCRIPTION OF KEY-WORD-IN-CONTEXT PROGRAM (KWIC)

The user wishing to produce a concordance will first choose his computer terminal and then his typing element. Any language that can be keyboarded from a typewriter and for which a typing element exists can be computer processed in upper and lower case with the respective diacritical marks. The print-out will be in perfect alphabetical order. The typing element may give some problem. The Portuguese element is IBM BR-971 which unfortunately has no number one separated from lower case L (l). The problem is resolved by a special output program which converts all number ones to lower case L. However, it is necessary to input the symbol = from the 2741 IBM terminal. Also, the question mark (?) presents a problem since on the BR-971 element it is on upper case 6 or key 19 which on the 2741 is the control input. The problem is resolved by first typing the question mark (?) and then the nul character, upper case key 41 or + on element BR-971. The Brazilian or Portuguese element will work equally well for French. Most of the elements for French are twelve pitch and they also have ¨ on the nul key position. For Spanish, the Puerto Rican element, No. 040 or part no. 1167040, presents no problems. There is a separate number one with the reverse question and exclamation marks. The Bulgarian typing element has been used for Russian, but it is understood that IBM has recently come out with three new elements with the Cyrillic alphabet.

With the special typing element the user prepares a file named after the language, using only the first four letters in caps. This file has three lines exactly. The lines are not sequenced-numbered. The first line contains the capital or upper-case letters of the alphabet in the desired order for sorting. The lower-case letters are on the second line in their sort-order. The third line (no spaces between the lines) has those diacritical marks which require the use of the back-space before typing them. The accent marks are also arranged in "dictionary" order. Also, symbols used for sorts are placed in this line and they may or may not appear in the print-out.

Next a languages file is prepared on the same terminal using the same typing element. Any name up to eight characters may be used - the computer program will ask for its name. The lines of the file are sequenced-numbered. Each new source is labeled on a line preceding the input material. Each label has eight characters, namely, three digits, two letters, and three digits. This label provides for identifying the source, its classification, and the page of computer input for later reference. In the input material of this file, the sequence for separable diacritical marks is letter, backspace, and diacritical mark.

Blanks have meaning. Each sequence-number must be followed by at least one blank. Capitalized words which are not proper names are to be preceded by at least two spaces, to show that the capitalization arises from being at the start of a sentence, or the like. Proper names at the beginning of sentences will have only one space before them, though this is usually wrong style.

In addition to the input file, two others will be needed. An intermediate-results file may be called KWIC or any other name. A file for storing final results will also be necessary. The contents of the KWIC file are of little importance since it gets scratched and changed.

Next the user chooses his L and R positions. These are the number of characters of context to the left and respectively to the right of the index point of the character of the sort field or the KEY word. The total of values of L and R can not exceed 112. In the programs FOR-KWIC, FOR-LIST, REV-KWIC, and REV-LIST L has been set at fifty-five and R at fifty-seven in the first two statements. The user should change these to suit his own terminal and taste. L need not be equal to R, but FOR-KWIC must agree with FOR-LIST, and REV-KWIC must agree with REV-LIST.

To begin the concordance program the user calls for OLD FOR-KWIC and keyboards RUN. The computer will ask for the input file. For the languages only the first four letters are to be written. Any intermediate file may be used. The intermediate file must next be sorted and this is easily done by running the program SHELGAME. Finally, the user calls for OLD FOR-LISTS and types RUN. The desired concordance can be obtained by going to the storage file. All this is assuming that a forward concordance is desired. For a backward concordance use REV-KWIC and REV-LIST.

The programs FOR-LIST and REV-LIST assume that the intermediate file is called KWIC. If some other name is used, the user will, of course, make this change in his file statement.

SUBPROGRAMS FOR ARRANGING LETTERS IN ALPHABETICAL ORDER

The alphabetical order for each language is used, except for the ch,ll, and rr of Spanish which uses a special output program. The program name is the first four letters of each language. Extra features, such as tags, can be added to the third or diacritical mark line for special sorts. The final output program can convert them to blanks.

```
FREN    ABCÇDEFGHIJKLMNOPQRSTUVWXYZ
        abcçdefghijklmnopqrstuvwxyz
        ---

PORT    ABCÇDEFGHIJKLMNOPQRSTUVWXYZ
        abcçdefghijklmnopqrstuvwxyz
        ----&*

RUSS    абвгдежзийклмнопрстуфхцчшщъыьэюя

SPAN    ABCDEFGHIJKLMNÑOPQRSTUVWXYZ
        abcdefghijklmnñopqrstuvwxyz
        -*
```

SAMPLE TELETYPE INPUT FORMAT

SPOKEN

100 018MS783
110 Eu preciso saber ainda se você recebeu a minha encomenda, a bôlsa com as
120 encomendas. Você recebeu?
130 017FS783
140 Ah, recebi sim. E diga à Zilá Maria que a moça perdeu o resto da encomenda
150 com o remédio da mamãe, tudo mais. Só só consegui pegar a bôlsa com as minhas coisas.
160 O resto ela perdeu.
170 018MS783
180 Sim, mas eh ... dentro da bôlsa aí iam duas cartas minhas sendo que uma tinha
190 dinheiro. Você achou?
200 017FS783
210 Ah, recebi sim e já paguei ao Coronel Taveiro e ela já me deu o trôco. Depois
220 eu mando para você quanto custou, numa carta no avião para ir mais seguro e o resto do
230 dinheiro você mandou guardar aqui, né?
240 018MS783
250 É sim. O resto do dinheiro é para ficar guardado com você. Depois eu explico
260 melhor por carta, está bem Então um beijão bem grandão pra você. E agora neste instante
270 eu estou escrevendo mais uma carta pra você explicando uma porção de coisas. Essa semana
280 eu tive provas quase todos os dias e estou trabalhando muito lá na fábrica, entendeu?
290 017FS783
300 Entendi, sim, meu amor, entendi sim. Mas faça o possível pra escrever porque eu
310 fico aqui com muitas saudades suas, né? E eu quero saber notícias.
320 018MS783
330 Tá bem. Eu vou escrever o máximo que eu puder, meu amor. Um beijão pra você bem
340 grandão.
350 017FS783
360 Um beijão bem grandão pra você também.
370 192FH819
380 Mamãe, é Norma. Tudo bem aqui. Espero que aí também. Recebi a carta da Lia
390 hoje e tou telefonando pra dar um grande abraço na senhora, muitas felicidades, saúde e
400 tudo de bom. As crianças não falam porque estão na piscina. Mais tarde é que êles vão
410 chegar. Êles mandam um grande beijo pra senhora, Olívia, Paulinho, nós todos, ouviu?

VERIFY

```
100 REM WRITTEN BY HAROLD KAPLAN
110 REM
120 REM DESCRIPTION - THIS PROGRAM WILL COMPARE TWO FILES
130 REM               AND WILL PRINT OUT DISCREPANCIES
140 REM               FOUND.  THIS IS TO CHECK ACCURACY OF
150 REM               INPUT DATA.
160 REM
170 REM - - - - - - - - - MAIN PROGRAM - - - - - - - - - -
180 PRINT "FIRST FILE";
190 INPUT A$
200 PRINT "SECOND FILE";
210 INPUT B$
220 FILE #1:A$
230 FILE #2:B$
240 LINPUT #1:X$
250 LINPUT #2:Y$
260 IF X$ = Y$ THEN 300
270 PRINT "DISCREPANCY"
280 PRINT X$
290 PRINT Y$
300 IF MORE #1 THEN 340
310 IF MORE #2 THEN 370
320 PRINT "ALL DONE"
330 STOP
340 IF MORE #2 THEN 240
350 PRINT A$;" IS LONGER"
360 STOP
370 PRINT B$;" IS LONGER"
380 STOP
390 END
```

FOR-KWIC

```
100 REM PROGRAMMED BY PROF. HAROLD KAPLAN, U.S. NAVAL ACADEMY
110 REM  THIS MAKES THE CONTEXTS FOR FORWARD CONCORDANCES.  IT
120 REM CALLS FKEY AND FCAPS.  DON'T FORGET TO CHANGE L AND R
130 REM HERE AND IN FOR-LIST.
140 LET L=55
150 LET R=57
160 LET L9$=""
170 FOR J=1 TO L
180 LET L9$=L9$&" "
190 NEXT J
200 LET R9$=""
210 FOR J=1 TO R
220 LET R9$=R9$&" "
230 NEXT J
240 PRINT "NAME OF INPUT FILE";
250 INPUT A$
260 PRINT "NAME OF LANGUAGE (First four letters only)"
270 INPUT B$
280 PRINT "In which file are the results to be stored"
290 INPUT C$
300 FILE #1:A$
310 FILE #2:B$
320 FILE #3:C$
330 SCRATCH #3
340 MARGIN #3:4095
350 LINPUT #2:B2$
360 CHANGE B2$ TO B
370 DIM B(100)
380 FOR J=1 TO B(0)
390 LET V(B(J))=1
400 LET W(B(J))=J
410 LET T(B(J))=1
420 NEXT J
430 DIM V(128),W(128),T(128)
440 LINPUT #2:B2$
450 CHANGE B2$ TO B
460 FOR J=1 TO B(0)
470 LET V(B(J))=1
480 LET W(B(J))=J
490 NEXT J
500 LINPUT #2:B2$
510 CHANGE B2$ TO B
520 FOR J=1 TO B(0)
530 LET D(B(J))=J
540 NEXT J
550 DIM D(128)
560 IF MORE #1 THEN 590
570 PRINT "EMPTY INPUT FILE"
580 STOP
590 LINPUT #1:L2$
```

FOR-KWIC (continued)

```
600 LET L$=FNC$(L2$)
610 IF FNL$(L$)="T" THEN 640
620 PRINT "FIRST LINE IS NOT LABEL:";L2$
630 STOP
640 LET L3$=L$
650 LET L4$=L3$
660 IF LEN(L4$).0 THEN 690
670 PRINT "DONE.  Now RUN SHELGAME"
680 STOP
690 LET C$=""
700 IF END #1 THEN 780
710 LINPUT #1:L2$
720 LET L$=FNC$(L2$)
730 IF FNL$(L$)="T" THEN 760
740 LET C$=C$&" "&L$
750 GO TO 700
760 LET L3$=L$
770 GO TO 800
780 LET L3$=""
790 GO TO 800
800 CHANGE L9$&C$&R9$ TO Z
810 REM FOR REVERSE CONCORDANCE FOR J=1 TO Z(0)/2
820 REM FOR REVERSE CONCORDANCE LET T9=Z(J)
830 REM FOR REVERSE CONCORDANCE LET Z(J)=Z(Z(0)+1-J)
840 REM FOR REVERSE CONCORDANCE LET Z(Z(0)+1-J)=T9
850 REM FOR REVERSE CONCORDANCE NEXT J
860 REM FOR REVERSE CONCORDANCE FOR J=2 TO Z(0)-1
870 REM FOR REVERSE CONCORDANCE IF Z(J),.8 THEN 830
880 REM FOR REVERSE CONCORDANCE LET T9=Z(J+1)
890 REM FOR REVERSE CONCORDANCE LET Z(J+1)=Z(J-1)
900 REM FOR REVERSE CONCORDANCE LET Z(J-1)=T9
910 REM FOR REVERSE CONCORDANCE NEXT J
920 FOR J=1 TO Z(0)
930 LET Y(J)=ASC(B)+V(Z(J))*(ASC(N)-ASC(B))
940 NEXT J
950 LET Y(0)=Z(0)
960 CHANGE Z TO Z$
970 CHANGE Y TO Y$
980 DIM Z(4095),Y(4095)
990 LET P=0
1000 LET P=POS(Y$,"N",P)
1010 IF P=0 THEN 1110
1020 CALL "FKEY":Z(),W(),T(),D(),P,K$,Q
1030 LIBRARY "FKEY","FCAPS"
1040 PRINT #3:K$&L4$&SEG$(Z$,P-L,P+R)&STR$(Q)
1050 LET P=POS(Y$,"B",P)
1060 IF P=0 THEN 1110
1070 IF Z(P),.8 THEN 1100
1080 LET P=P+2
1090 GO TO 1050
```

FOR-KWIC (continued)

```
1100 GO TO 1000
1110 GO TO 650
1120 DEF FNC$(A$)
1130 LET A=POS(A$," ",1)
1140 REM LET A=0
1150 LET FNC$=SEG$(A$,A+1,LEN(A$))
1160 FNEND
1170 DEF FNL$(A$)
1180 IF LEN(A$),.8 THEN 1260
1190 LET A=POS(A$," ",1)
1200 IF A.0 THEN 1260
1210 FOR J=1 TO 8
1220 IF SEG$(A$,J,J).=CHR$(97) THEN 1260
1230 NEXT J
1240 LET FNL$="T"
1250 GO TO 1270
1260 LET FNL$="F"
1270 FNEND
1280 END
```

FKEY

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  THIS GENERATES THE SORT-KEYS FOR FORWARD CONCORDANCES.
120 REM K$ IS THE KEY, AND Q K$ IS THE LENGTH OF THE WORD FOUND.
130 REM FKEY IS CALLED BY FOR-KWIC.
140 SUB "FKEY":Z(),W(),T(),D(),P,K$,Q
150 CALL "FCAPS":Z(),P,(ASC(a)-ASC(A))
160 LET F$="AAAA
```

FCAPS

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  THIS CHECKS FOR CAPITALS AT THE BEGINNINGS OF SENTENCES
120 REM BY LOOKING FOR THE SPACES BEFORE.  IT IS NEEDED BY FOR-KWIC
130 REM AND FOR-LIST.
140 SUB "FCAPS":Z(),P,D
150 IF Z(P-2),.ASC( ) THEN 180
160 IF Z(P-1),.ASC( ) THEN 180
170 LET Z(P)=Z(P)+D
180 SUBEND
```

SHELGAME

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  THIS RATHER TRIVIAL PROGRAM MERELY READS IN "KWIC", CALLS
120 REM FOR A SORT, SCRATCHES "KWIC", AND WRITES OUT THE RESULT
130 REM INTO "KWIC".
140 FILE #1:"KWIC"
150 DIM X$(3000)
160 LET J=J+1
170 LINPUT #1:X$(J)
180 IF MORE #1 THEN 160
190 CALL "SHELSORT":X$(),J
200 LIBRARY "SHELSORT"
210 SCRATCH #1
220 MARGIN #1:4000
230 FOR K=1 TO J
240 PRINT #1:X$(K)
250 NEXT K
260 PRINT J;"ITEMS SORTED.  Now go to FOR-LIST or REV-LIST"
270 END
```

SHELSORT

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  THIS IS THE FAMOUS SORT BY SHELL'S ALGORITHM.  N IS HOW MAN
120 REM STRINGS THERE ARE IN THE ARRAY X$().  THEY ARE SORTED IN
130 REM ASCENDING ORDER.
140 SUB "SHELSORT":X$(),N
150 LET H=INT(N/2)
160 IF H=0 THEN 400
170 FOR W=1 TO H
180 LET T=-1
190 FOR J=W TO N-H STEP H
200 IF X$(J),=X$(J+H) THEN 250
210 LET T=+1
220 LET Q$=X$(J)
230 LET X$(J)=X$(J+H)
240 LET X$(J+H)=Q$
250 NEXT J
260 IF T,0 THEN 370
270 LET T=-1
280 FOR J=W+H*INT((N-W)/H) TO 1+H STEP -H
290 IF X$(J-H),=X$(J) THEN 340
300 LET T=+1
310 LET Q$=X$(J)
320 LET X$(J)=X$(J-H)
330 LET X$(J-H)=Q$
340 NEXT J
350 IF T,0 THEN 370
360 GO TO 180
370 NEXT W
380 LET H=INT(H/2)
390 GO TO 160
400 SUBEND
```

FOR-LIST

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  This reads in "KWIC", reformats, and prints
120 REM  the results.  It calls FCAPS.  Don't forget to
130 REM  change L and R to agree with FOR-KWIC.
140 PRINT "In which file are the results to be stored"
150 INPUT S$
160 FILE #2:S$
170 SCRATCH #2
180 MARGIN #2:130
190 LET L=55
200 LET R=57
210 LIBRARY "FCAPS"
220 MARGIN 4000
230 FILE #1:"KWIC"
240 LINPUT #1:X$
250 LET A$=SEG$(X$,91,91+8-1)
260 LET B$=SEG$(X$,91+8,91+8+L-1)
270 LET C$=SEG$(X$,91+8+L,91+8+L+1+R-1)
280 LET D$=SEG$(X$,91+8+L+1+R,LEN(X$))
290 LET D=VAL(D$)
300 LET E$=SEG$(B$&C$,L-1,L+D)
310 CHANGE E$ TO Z
320 DIM Z(290)
330 CALL "FCAPS":Z(),3,(ASC(a)-ASC(A))
340 CHANGE Z TO E$
350 LET E$=SEG$(E$,3,LEN(E$))
360 IF E$=E2$ THEN 430
370 IF F,. 1 THEN 400
380 PRINT #2
390 GO TO 410
400 PRINT #2: TAB(L);F
410 LET F=0
420 LET E2$=E$
430 LET P=0
440 LET H$=""
450 LET P=POS(B$,CHR$(8),P+1)
460 IF P=0 THEN 490
470 LET H$=H$&"  "
480 GO TO 450
490 LET B$=H$&B$
500 PRINT #2: B$;TAB(L+1);C$;TAB(L+R+3);A$
510 LET F=F+1
520 IF MORE #1 THEN 240
530 PRINT #2: TAB(L);F
540 END
```

REV-KWIC

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  THIS IS THE CONTEXT-MAKER FOR REVERSE CONCORDANCES.
120 REM DON'T FORGET TO CHANGE R AND L HERE AND IN REV-LIST.
130 LET L=40
140 LET R=40
150 LET L9$=""
160 FOR J=1 TO L
170 LET L9$=L9$&" "
180 NEXT J
190 LET R9$=""
200 FOR J=1 TO R
210 LET R9$=R9$&" "
220 NEXT J
230 PRINT "Input file, Language, Output file"
240 INPUT A$,B$,C$
250 FILE #1:A$
260 FILE #2:B$
270 FILE #3:C$
280 SCRATCH #3
290 MARGIN #3:4095
300 LINPUT #2:B2$
310 CHANGE B2$ TO B
320 DIM B(100)
330 FOR J=1 TO B(0)
340 LET V(B(J))=1
350 LET W(B(J))=J
360 LET T(B(J))=1
370 NEXT J
380 DIM V(128),W(128),T(128)
390 LINPUT #2:B2$
400 CHANGE B2$ TO B
410 FOR J=1 TO B(0)
420 LET V(B(J))=1
430 LET W(B(J))=J
440 NEXT J
450 LINPUT #2:B2$
460 CHANGE B2$ TO B
470 FOR J=1 TO B(0)
480 LET D(B(J))=J
490 NEXT J
500 DIM D(128)
510 IF MORE #1 THEN 540
520 PRINT "EMPTY INPUT FILE"
530 STOP
540 LINPUT #1:L2$
550 LET L$=FNC$(L2$)
560 IF FNL$(L$)="T" THEN 590
570 PRINT "FIRST LINE IS NOT LABEL:";L2$
580 STOP
590 LET L3$=L$
```

REV-KWIC (continued)

```
600 LET L4$=L3$
610 IF LEN(L4$).0 THEN 640
620 PRINT "DONE - Now go to SHELGAME"
630 STOP
640 LET C$=""
650 IF END #1 THEN 730
660 LINPUT #1:L2$
670 LET L$=FNC$(L2$)
680 IF FNL$(L$)="T" THEN 710
690 LET C$=C$&" "&L$
700 GO TO 650
710 LET L3$=L$
720 GO TO 750
730 LET L3$=""
740 GO TO 750
750 CHANGE L9$&C$&R9$ TO Z
760 FOR J=1 TO Z(0)/2
770 LET T9=Z(J)
780 LET Z(J)=Z(Z(0)+1-J)
790 LET Z(Z(0)+1-J)=T9
800 NEXT J
810 FOR J=2 TO Z(0)-1
820 IF Z(J),.8 THEN 860
830 LET T9=Z(J+1)
840 LET Z(J+1)=Z(J-1)
850 LET Z(J-1)=T9
860 NEXT J
870 FOR J=1 TO Z(0)
880 LET Y(J)=ASC(B)+V(Z(J))*(ASC(N)-ASC(B))
890 NEXT J
900 LET Y(0)=Z(0)
910 CHANGE Z TO Z$
920 CHANGE Y TO Y$
930 DIM Z(4095),Y(4095)
940 LET P=0
950 LET P=POS(Y$,"N",P)
960 IF P=0 THEN 1060
970 CALL "RKEY":Z(),W(),T(),D(),P,K$,Q
980 LIBRARY "RKEY","RCAPS"
990 PRINT #3:K$&L4$&SEG$(Z$,P-L,P+R)&STR$(Q)
1000 LET P=POS(Y$,"B",P)
1010 IF P=0 THEN 1060
1020 IF Z(P),.8 THEN 1050
1030 LET P=P+2
1040 GO TO 1000
1050 GO TO 950
1060 GO TO 600
1070 DEF FNC$(A$)
1080 LET A=POS(A$," ",1)
1090 REM LET A=0
```

REV-KWIC (continued)

```
1100 LET FNC$=SEG$(A$,A+1,LEN(A$))
1110 FNEND
1120 DEF FNL$(A$)
1130 IF LEN(A$),.8 THEN 1210
1140 LET A=POS(A$," ",1)
1150 IF A.0 THEN 1210
1160 FOR J=1 TO 8
1170 IF SEG$(A$,J,J).=CHR$(97) THEN 1210
1180 NEXT J
1190 LET FNL$="T"
1200 GO TO 1220
1210 LET FNL$="F"
1220 FNEND
1230 END
```

RKEY

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  THIS IS THE SORT-KEY MAKER FOR REVERSE CONCORDANCES.
120 REM  IT IS CALLED BY REV-KWIC.  K$ IS THE RESULTING KEY,
130 REM  AND Q IS THE LENGTH OF THE WORD FOUND.
140 SUB "RKEY":Z(),W(),T(),D(),P,K$,Q
150 CALL "RCAPS":Z(),P,(ASC(a)-ASC(A))
160 LET F$="AAAA
```

RCAPS

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  This is needed by REV-KWIC and REV-LIST.  It looks
120 REM  for starts of sentences marked by two spaces, and
121 REM  changes capitals to lower-case.
130 SUB "RCAPS":Z(),P,D
140 IF Z(P+2),.ASC( ) THEN 170
150 IF Z(P+1),.ASC( ) THEN 170
160 LET Z(P)=Z(P)+D
170 SUBEND
```

REV-LIST

```
100 REM PROGRAMMED BY HAROLD KAPLAN
110 REM  THIS REFORMATS "KWIC" FOR REVERSE CONCORDANCES AND PRINTS
120 REM THE RESULT.  DON'T FORGET TO CHANGE L AND R HERE TO AGREE
130 REM WITH REV-KWIC.
140 PRINT  "In which file are the results to be stored"
150 INPUT S$
160 FILE #2:S$
170 SCRATCH #2
180 MARGIN #2:130
190 LET L=40
200 LET R=40
210 LIBRARY "FCAPS"  'YES IT IS FCAPS, NOT RCAPS.
220 MARGIN 4000
230 FILE #1:"KWIC"
240 LINPUT #1:X$
250 LET A$=SEG$(X$,91,91+8-1)
260 LET G$=SEG$(X$,91+8,91+8+L+1+R-1)
270 CHANGE G$ TO G
280 FOR J=1 TO G(0)/2
290 LET T9=G(J)
300 LET G(J)=G(G(0)+1-J)
310 LET G(G(0)+1-J)=T9
320 NEXT J
330 CHANGE G TO G$
340 LET B$=SEG$(G$,1,R+1)
350 LET C$=SEG$(G$,R+2,R+L+1)
360 DIM G(300)
370 LET D$=SEG$(X$,91+8+L+1+R,LEN(X$))
380 LET D=VAL(D$)
390 LET E$=SEG$(B$,LEN(B$)-D+1-2,LEN(B$))
400 CHANGE E$ TO Z
410 DIM Z(290)
420 CALL "FCAPS":Z(),3,(ASC(a)-ASC(A))
430 CHANGE Z TO E$
440 LET E$=SEG$(E$,3,LEN(E$))
450 IF E$=E2$ THEN 520
460 IF F,.1 THEN 490
470 PRINT #2
480 GO TO 500
490 PRINT #2: TAB(L-1);F
500 LET F=0
510 LET E2$=E$
520 LET P=0
530 LET H$=""
540 LET P=POS(B$,CHR$(8),P+1)
550 IF P=0 THEN 580
560 LET H$=H$&"  "
570 GO TO 540
580 LET B$=H$&B$
590 PRINT #2: B$;TAB(L+2);C$;TAB(L+R+4);A$
```

PRINT-OUT KWIC (Reduced size)

também. A Natali estêve aqui com	a Ana Cristina, a Lea, a Judith, êsse	590FH819
se você recebeu a minha encomenda,	a bôlsa com as encomendas. Você re	018MS783
tudo mais. Só só consegui pegar	a bôlsa com as minhas coisas. O resto	017FS783
i tudo bem, e eu vou (escr- responder	a carta da Lia amanhã ou depois.	192FH819
i. Espero que aí também. Recebi	a carta da Lia hoje e tou telefonando p	192FH819
o, chega dia vinte. E você aguarde	a chegada dela, viu?	590FH819
mais mate pra você. E outra coisa,	a Ítala embarcou no sábado. (Deve	590FH819
êve aqui com a Ana Cristina, a Lea,	a Judith, êsse pessoal, sabe - Agora	590FH819
ali estêve aqui com a Ana Cristina,	a Lea, a Judith, êsse pessoal, sabe -	590FH819
a tou esperando o pessoal à noite.	A Marli já chegou aí - Com - com o	590FH819
Todos bem aqui de saúde também.	A Natali estêve aqui com a Ana Cristi	590FH819
preciso saber ainda se você recebeu	a minha encomenda, a bôlsa com as enc	018MS783
ebi sim. E diga à Zilá Maria que	a moça perdeu o resto da encomenda com 13	017FS783
sabe - Agora tou esperando o pessoal	à noite. A Marli já chegou aí -	590FH819
a fora, na praia, voltamos no domingo	à noite, aqui vai tudo bem, e eu vou	192FH819
Ah, recebi sim. E diga	à Zilá Maria que a moça perdeu o re 3	017FS783
r em Nova Iorque lá pro dia dezoito	a& vinte de setembro. Se ela nun chega	590FH819
e e tou telefonando pra dar um grande	abraço na senhora, muitas felicidades,	192FH819
sendo que uma tinha dinheiro. Você	achou?	018MS783
beijão bem grandão pra você. E	agora neste instante eu estou escrevend	018MS783
ea, a Judith, êsse pessoal, sabe -	Agora tou esperando o pessoal à noite 2	590FH819
a dezoito, chega dia vinte. E você	aguarde a chegada dela, viu?	590FH819
	Ah, recebi sim. E diga à Zilá Mari	017FS783
	Ah, recebi sim e já paguei ao Coronel 2	017FS783
ssoal à noite. A Marli já chegou	aí - Com - com o mate que eu mandei,	590FH819

RCONCORD REVERSE CONCORDANCE OUTPUT (Reduced size)

ora tou esperando o pessoal à noite.	A	Marli já chegou aí - Com - com o m	590FH819
. Todos bem aqui de saúde também.	A	Natali estêve aqui com a Ana Cristina	590FH819
da se você recebeu a minha encomenda,	a	bôlsa com as encomendas. Você rece	018MS783
stêve aqui com a Ana Cristina, a Lea,	a	Judith, êsse pessoal, sabe - Agora t	590FH819
atali estêve aqui com a Ana Cristina,	a	Lea, a Judith, êsse pessoal, sabe -	590FH819
, mais mate pra você. E outra coisa,	a	Ítala embarcou no sábado. (Deve E	590FH819
ito, chega dia vinte. E você aguarde	a	chegada dela, viu?	590FH819
ecebi sim. E diga à Zilá Maria que	a	moça perdeu o resto da encomenda com o	017FS783
qui. Espero que aí também. Recebi	a	carta da Lia hoje e tou telefonando pra	192FH819
e também. A Natali estêve aqui com	a	Ana Cristina, a Lea, a Judith, êsse p	590FH819
e, tudo mais. Só só consegui pegar	a	bôlsa com as minhas coisas. O resto e	017FS783
vai tudo bem, e eu vou (escr- responder	a	carta da Lia amanhã ou depois.	192FH819
u preciso saber ainda se você recebeu	a	minha encomenda, a bôlsa com as encom	018MS783
13			
Ah, recebi sim. E diga	à	Zilá Maria que a moça perdeu o resto	017FS783
sabe - Agora tou esperando o pessoal	à	noite. A Marli já chegou aí - Com	590FH819
a fora, na praia, voltamos no domingo	à	noite, aqui vai tudo bem, e eu vou (esc	192FH819
3			
estou trabalhando muito lá na fábrica		, entendeu?	018MS783
di, sim, meu amor, entendi sim. Mas faça		o possível pra escrever porque eu fic	017FS783
sim. E diga à Zilá Maria que a moça		perdeu o resto da encomenda com o rem	017FS783
ero que aí também. Recebi a carta da		Lia hoje e tou telefonando pra dar um g	192FH819
bem, e eu vou (escr- responder a carta da		Lia amanhã ou depois.	192FH819
semana porque nós fomos passar o fim da		semana fora, na praia, voltamos no domi	192FH819
... não ... nós eu nun pedi no fim da		semana porque nós fomos passar o fim	192FH819
u o resto da encomenda com o remédio da		mamãe, tudo mais. Só só consegui	017FS783
Sim, mas eh ... dentro da		bôlsa aí iam duas cartas minhas sen	018MS783
Zilá Maria que a moça perdeu o resto da		encomenda com o remédio da mamãe, t	017FS783